DOCUMENT RESUME

ED 222 512 TM 820 656

AUTHOR Wildemuth, Barbara M., Comp.
TITLE A Bibliography to Accompany the Joint Committee's Standards on Educational Evaluation.
INSTITUTION ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO ERIC-TM-81
PUB DATE Dec 81
CONTRACT 400-78-0003
NOTE 114p.
AVAILABLE FROM ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, Princeton, NJ 08541 ($8.50).

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Annotated Bibliographies; Cost Effectiveness; Ethics; *Evaluation Methods; Evaluators; *Literature Reviews; Measurement Techniques; Political Power; Research Methodology; *Research Utilization; *Standards
IDENTIFIERS Evaluation Utilization; *Standards for Evaluation Educ Prog Proj Materials

ABSTRACT

An annotated bibliography is organized to match the format of the Standards for Evaluations of Educational Programs, Projects, and Materials, including sections on utility, feasibility, propriety, and accuracy of evaluations. It identifies literature that includes in-depth information about the issues covered in the Standards. The Standards are intended to guide evaluations of programs, projects, or materials in elementary, secondary, higher, or adult education. The intended audience includes persons who commission, conduct, or use evaluations, especially teachers, administrators, evaluators, curriculum specialists, school board members, legislators, counselors, leaders of educational associations, and parents. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED222512

ERIC/TM Report 81

A BIBLIOGRAPHY TO ACCOMPANY THE JOINT COMMITTEE'S STANDARDS ON EDUCATIONAL EVALUATION

Compiled by

Barbara M. Wildemuth

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

December 1981

TM 820 452

ERIC Clearinghouse on Tests, Measurement, and Evaluation

Educational Testing Service, Princeton, New Jersey 08541

The material in this publication was prepared pursuant to a contract with the National Institute of Education, U.S. Department of Education. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to qualified professionals for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either these reviewers or the National Institute of Education.

ERIC Clearinghouse on Tests, Measurement,
and Evaluation
Educational Testing Service
Princeton, NJ 08441

December 1981

TABLE OF CONTENTS

The Joint Committee on Standards for Educational Evaluation i
Introduction ii

A. Utility Standards
- A1. Audience Identification 2
- A2. Evaluator Credibility 5
- A3. Information Scope and Selection 8
- A4. Valuational Interpretation 11
- A5. Report Clarity 20
- A6. Report Dissemination 23
- A7. Report Timeliness 25
- A8. Evaluation Impact 26

B. Feasibility Standards
- B1. Practical Procedures 34
- B2. Political Viability 36
- B3. Cost Effectiveness 41

C. Propriety Standards
- C1. Formal Obligation 43
- C2. Conflict of Interest 46
- C3. Full and Frank Disclosure 49
- C4. Public's Right to Know 50
- C5. Rights of Human Subjects 51
- C6. Human Interactions 54
- C7. Balanced Reporting 56
- C8. Fiscal Responsibility 57

D. Accuracy Standards
- D1. Object Identification 58
- D2. Context Analysis 61
- D3. Described Purposes and Procedures 63
- D4. Defensible Information Sources 64
- D5. Valid Measurement 67
- D6. Reliable Measurement 71
- D7. Systematic Data Control 74
- D8. Analysis of Quantitative Information 76
- D9. Analysis of Qualitative Information 80
- D10. Justified Conclusions 86
- D11. Objective Reporting 90

General Monographs and Textbooks 91

Index 103

THE JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION

Daniel L. Stufflebeam (Western Michigan University), Chairman

Larry Braskamp (University of Illinois), representing the Association of Institutional Researchers

Esther E. Diamond (Science Research Associates), representing the Association for Measurement and Evaluation in Guidance

Roy Forbes (Education Commission of the States), representing the Education Commission of the States

Freda Holley (Austin, Texas, Independent School District), representing the American Educational Research Association

Phillip Hosford (New Mexico State University), representing the Association for Supervision and Curriculum Development

Robert L. Linn (University of Illinois), representing the National Council on Measurement in Education

William Mays, Jr. (Michigan Elementary and Middle Schools Principals Association), representing the National Association for Elementary School Principals

Bernard McKenna (National Education Association), representing the National Education Association

Lloyd Nielsen (Roseville, Minnesota, School District), representing the American Association of School Administrators

James Oglesby (Columbia, Missouri, Public Schools), representing the National School Boards Association

James R. Sanders (Western Michigan University), representing the Evaluation Network

Carol Kehr Tittle (University of North Carolina at Greensboro), representing the American Psychological Association

INTRODUCTION

As a means of improving educational evaluations, 12 professional organizations[1] in 1975 appointed a 17 member joint committee and charged it with devising standards for educational evaluations. Following five years of development--including extensive review, field tests and hearings--the Standards for Evaluations of Educational Programs, Projects, and Materials was published.[2] The Standards are intended to guide evaluations of programs, projects or materials in elementary, secondary, higher, or adult education. The intended audience includes persons who commission, conduct, or use evaluations, especially teachers, administrators, evaluators, curriculum specialists, school board members, legislators, counselors, leaders of educational associations, and parents.

During the development of the Standards, the Joint Committee considered the suggestion that references be included. After a good deal of discussion, the Joint Committee decided against including references, because it was felt that the standards should stand on their own, that references could quickly become dated, and that references might be inappropriately taken as a view specifically endorsed by the Joint Committee. The idea of a separate bibliography specifically keyed to the Standards seemed a reasonable compromise that would help satisfy an expressed need for background reading on the topics covered by the Standards. This bibliography is intended to help users of the Standards to identify literature that includes in-depth information about the issues covered in each of the standards. It is also the first of a number of derivative documents and training aides that are being developed to supplement the Standards and enhance their utility.

---

[1] American Association of School Administrators, American Educational Research Association, American Federation of Teachers, American Personnel and Guidance Association, American Psychological Association, Association for Supervision and Curriculum Development, Council for American Private Education, Education Commission of the States, National Association of Elementary School Principals, National Council on Measurement in Education, National Education Association, National School Boards Association. The Joint Committee in 1981 reconstituted itself. The current Sponsoring Organizations and representatives are listed on the preceding page.

[2] Joint Committee on Standards for Educational Evaluation, Standards for Evaluations of Educational Programs, Projects, and Materials. New York: McGraw-Hill Book Company, 1981.

The ERIC Clearinghouse on Tests, Measurement, and Evaluation was pleased to develop this bibliography with the cooperation of the Joint Committee on Standards for Educational Evaluation, Inc. The effort to produce a bibliography began when Robert Carlson of the University of Vermont developed an initial draft for the Joint Committee while he was a visiting scholar at the Evaluation Center at Western Michigan University. Muriel Katzenmeyer, a research assistant at the Center, added entries to the Carlson draft. ERIC began its work on the bibliography with a computerized literature search of the ERIC database from 1966 through mid-1981. The references of the documents and articles identified by this search were checked for additional relevant citations. Finally, the entries from the ERIC draft were compared with those from the Carlson-Katzenmeyer draft, and non-overlapping citations were added. At this point, annotations were prepared for all items in the bibliography.

This annotated bibliography was submitted to the Joint Committee for review. Additions and deletions suggested by them were coordinated by Larry Braskamp and Carol Tittle, members of the Joint Committee, and Robert Rodosky, Staff Director for the Joint Committee. They were assisted by Paul Mayberry, a graduate student at the University of Illinois. These suggestions were incorporated in the final revision of the bibliography.

The main body of the bibliography is organized to match the Standards, chapter by chapter. Each entry begins with the descriptor for each standard (e.g., Evaluation Impact) and its definition. Following the descriptor and definition, the annotated references appear in alphabetical order by author. For those references in the Educational Resources Information Center (ERIC) system, the ERIC accession number is also listed. The Committee decided to include only the most current references, except for those they consider "classic."

Each reference is listed only once, and closely related sections are cross referenced. Citations referring to more than one of the standards are included in a separate section labeled "General Monographs and Textbooks." The final section of the publication is an author index.

We would like to thank all the people associated with the preparation of this bibliography: Robert Carlson and Muriel Katzenmeyer developed an early draft; Kathryn Hecht of the University of Alaska and Leonard Cahen of Arizona State University shared extensive evaluation course bibliographies; and James Sanders of the Evaluation Network provided additional course bibliographies that were obtained as a result of a survey by that organization. Past and present Joint Committee members helped by reviewing drafts and adding and deleting entries. In addition, the following members were particularly helpful: Larry Braskamp, Henry Brickell, Don Campbell, Ron Carver, Ester Diamond, Egon Guba, Robert Linn, George Madaus, Bernard McKenna, Lorrie Shepard, Daniel Stufflebeam and Carol Kehr Tittle.

We all hope that this bibliography will assist users of the Standards to develop in-depth knowledge of the underlying principles and ways of applying them.

December 1981

Barbara M. Wildemuth
Associate Director
ERIC Clearinghouse on Tests,
Measurement and Evaluation

Robert J. Rodosky
Staff Director
Joint Committee on Standards
for Educational Evaluation

A. UTILITY STANDARDS

1. Audience Identification. Audiences involved in or affected by the evaluation should be identified, so that their needs can be addressed.

Hess, Robert J.; Wright, William J. Evaluation Strategies as a Function of Product Development Stages. St. Ann, Mo.: Central Midwestern Regional Educational Lab., 1972. 30p. ED 064 364.

There are issues in curriculum evaluation and stages of product development that demand the use of experimental or quasi-experimental designs. To counteract criticism of evaluation efforts, an approach to the examination of the multiple issues involved in curriculum product evaluation across the usual developmental cycle of educational products was developed. Curriculum products typically move through a developmental sequence comprised of five stages: initial state, hot house (the initial tryout of a prototype product), pilot test, field test, and public diffusion. Each stage represents a milestone in the life of a product. In the course of evaluation, various audiences are acquired: the sponsor, the institution, the developer, consumer representatives, and advisors. There are five major dimensions of a comprehensive evaluation of curriculum products: desirability/feasibility, management/procedural cost, product worth, usability, and generalizability. Issues relating to the continuation or termination of a program concern statement and fulfillment of objectives, establishing a rationale for the use of particular measuring instruments, determination of whether or not different effects result from alternative procedures. When the product enters the diffusion stage, formative evaluation is ended and summative evaluation should begin. It is pointed out that true summative evaluation is consumer protection and is a three tiered process, wherein: (1) the product developer establishes the criteria; (2) some agency of the federal government examines the product; and (3) local education agencies research the products.

House, Ernest R. The Logic of Evaluative Argument. CSE Monograph Series in Evaluation, 7. Los Angeles: Center for the Study of Evaluation, Univ. of California, 1977. ED 156 719.

Evaluation is an act of persuasion directed to a specific audience concerning the solution of a problem. The process of evaluation is prescribed by the nature of knowledge--which is generally complex, always uncertain (in varying degrees), and not always propositional--and by the nature of logic, which is always selective. In the process of persuasion one must ascertain who the audience is and find a basis of agreement on premises, both of facts and values, and on presumptions. Two criteria for evaluation are: the most efficient way to a given end, or the most effective use of available resources. Quantitative evaluation methods involve three stages: (1) substantive definition of the problem and its translation into a formal, mathematical model; (2) compilation of information in terms of the formal model and its formal, logical analysis; and (3) translation of the formal conclusions back into substantive terms. Both formulation and interpretation require good intuitive judgment. The evaluator and the audience must employ their reasoning in a dialogue, and both must assume responsibility, since evaluation is never completely convincing nor entirely arbitrary. The logical arguments used in two works are discussed. The works--Gene V. Glass' "Evaluation Skills," and Scriven's reply--are appended.

McGranahan, Pamela. Implications of Client Demands for R & E Activities. Unpublished. 13p. ED 167 592.

Potential clients for centralized school district organizations with societies, federal and state governments, boards of education, superintendents, other administrators, principals, and teachers. A historical review of the evaluation literature supports the proposition that some research and evaluation Unit (R & E) clients are served more directly than others; that this service to particular clients is in response to their demands; and that service to all clients is shaped by the demands of the most direct clients. Historically, superintendents were the most directly served clients; in responding to their administrative needs, R & E units engaged primarily in data collection activities. Despite federal and state evaluation requirements mandated by the Elementary and Secondary Education ACT, R & E units may still be engaged primarily in data collection activities (such as the reporting of norm referenced test scores) rather than in evaluation activities. It is likely that instructional clients such as teachers and project directors will be less directly served than administrative clients.

Patton, Michael Q. The Personal Factor: Identification and Organization of Relevant Decision-Makers and Information-Users. In Patton, Michael Q., Utilization-Focused Evaluation. Beverly Hills, Calif.: SAGE Pub., 1978. Chapter 4.

The first step in the utilization-focused approach to evaluation is identification and organization of relevant decision-makers for and information-users of the evaluation. In a study of factors affecting evaluation utilization, two factors emerged with consistency: the political considerations factor and the personal factor. The personal factor refers to the presence of an identifiable individual or group of people who personally cared about the evaluation and the information it generated. Identification of these relevant decision-makers and information-users, and determination of their information needs are critical the utilization of the evaluation. Evaluators frequently avoid this identification process by themselves becoming the major decision-makers for the evaluation, by using the standard "identification of audience" approach, by focusing on the decisions and information rather than the decision-makers and information-users, by deciding that the funders of the evaluation and/or program are the relevant information-users, or by targeting evaluations at organizations rather than at individuals.

Straton, Ralph G. Ethical Issues in Evaluating Educational Programs. Studies in Educational Evaluation, v3 n1 p57-66, Spring 1977. EJ 180 463.

Five broad ethical issues which face evaluators in the conduct of evaluation studies are discussed: (1) the identification of the audiences to be served by the evaluation study; (2) the choice of variables to be examined and the sources of information to be used in the study; (3) the technical adequacy and cost-effectiveness of the instruments and procedures to be employed; (4) the rights to privacy and confidentiality of subjects and program personnel; and (5) the relationships between the evaluator and program sponsors, participants and audiences.

A. UTILITY STANDARDS

2. Evaluator Credibility. The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that their findings achieve maximum credibility and acceptance.

Ahn, Unhai R.; And Others. Spectrum of Objectivity-Credibility in Evaluation. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C., 1975. 17p. ED 106 367.

Evaluation roles used in the Department of Research and Development in the Cincinnati Public Schools are identified and described. These include: project evaluator, local-school evaluator, independent-program evaluator, external evaluator and external auditors. The merits of each evaluation role are discussed as to its relationship with credibility, objectivity, independence and usefulness. The basis for judging the merits of each evaluation role with regard to the above four criteria are: (1) types of decisions to be made; and (2) safeguards to maximize each of the four criteria.

Braskamp, Larry A.; And Others. The Credibility of a Local Educational Program Evaluation Report: Author Source and Client Audience Characteristics. American Educational Research Journal, v15 n3 p441-450, Summer 1978.

The judged usefulness and objectivity of a simulated evaluation report and client agreement with the report's recommendations were examined as functions of the evaluator's simulated professional background, e.g., "researcher," "evaluator," or "art educator," and the client's organizational role status (teachers or administrators). The results suggest that source and audience characteristics influence client ratings of the evaluator but do not effect changes in agreement with the evaluator's recommendations.

Gurel, Lee. The Human Side of Evaluating Human Services Programs: Problems and Prospects. In Guttentag, Marcia; Struening, Elmer L. (Eds.), Handbook of Evaluation Research. Volume 2. Beverly Hills: SAGE Pub., 1975. Chapter 2.

The thesis presented here is that organizational context, structural constraints and requirements, and interpersonal relationships have profound consequences for the success or failure of evaluative activity. Four considerations related to the context within which the manager and evaluator interact are examined: (1) the conflicting superordinate organizational goals to which the manager and evaluator subscribe; (2) the stereotype of scientific omnipotence; (3) the extension of rigorous evaluation to areas of public service only recently considered exempt from external scrutiny; and (4) the recourse to evaluation as a pancreas for programs in failing health. Within this context, four areas of manager/evaluator interaction are potential sources of friction: (1) identifying program objectives, rationale, and procedures; (2) differing motivations for evaluation; (3) demands on the operating staff; and (4) the use of rigorous evaluation designs.

Millman, Jason. Selecting Educational Researchers and Evaluators. TM Report 48. Princeton, NJ.: ERIC Clearinghouse on Tests, Measurement, and Evaluation, December 1975. 15p. ED 117 191.

Aimed at those individuals who are in a position to hire or promote educational researchers or evaluators, this paper provides some practical suggestions for assessing these personnel. Selection of a research or evaluation (R & E) firm is not treated separately from the task of hiring an individual; the quality of work done by a firm depends largely on the people who do the job. Much consideration should be given to specifying job descriptions and requirements. The value of R & E competencies depends upon the specific tasks expected to be performed. A synthesis of the efforts of a task force of the American Educational Research Association to identify educational R & E competencies groups under 25 general tasks is included in the document. In an effort to identify a universe of evaluation competencies, Stufflebeam and Bunda produced approximately 250 items groups under eight major categories. The categories and examples of corresponding self-assessment items are also included. Several strategies for assessing whether an individual possesses the competencies needed for a specific job are considered. These include discussions of certification, formal training, testing, R & E output, bibliographic and academic characteristics, and membership in special professional associations and directories having more stringent entry requirements than presently exist.

Newman, Dianna L.; And Others. Communication Theory and the Utilization of Evaluation. New Directions for Program Evaluation, n5 p29-35, 1980. EJ 229 191.

The process of reporting evaluation results is described in terms of communications theory. Results of several simulated studies of factors which influence the credibility of evaluators and reports are summarized. The implications of the results cited are that: (1) the evaluator's credibility can be affected by the evaluator's title, sex, or source of information; (2) credibility and perception of the evaluation are affected by use of jargon and data, and type of evaluative information presented; and (3) the credibility of both evaluator and report can be affected by the receiver's organizational position, professional level and field, and perceived need for evaluation.

Newman, Warren B. Desirable Qualifications for Personnel Conducting Educational Program Evaluations and Audits. Paper presented at the annual meeting of the American Educational Research Association, 1976. 13p. ED 128 389.

A study was made of professional qualifications for personnel employed as program evaluators and auditors. These qualifications, according to operational or theoretical models, are necessary to assure local school districts of obtaining the services of competent and ethical personnel. Findings of: (1) a review of the literature; (2) a national survey of directors and staffs of research and program evaluation departments of public schools; (3) a review of representative contractual relations and job qualifications in use; (4) a survey of ten university training programs; and (5) a survey of legislators to determine the attributes of an evaluation report which make it acceptable as a basis for decision making, are reported. Criteria for employment of program evaluators and auditors are recommended, and the political implications of an accrediting process are discussed.

Sanders, James R. School Professionals and the Evaluation Function. Journal of School Psychology, v16 n4 p301-311, Winter 1978.

Evaluation is assumed to be an integral part of the professional delivery of school services. As such, professionals employed in school systems are called upon to define alternative roles they might play in evaluation, to consider alternative ways to organize for evaluation, and to focus on various objects of evaluation. Listed alternatives were drawn from emerging literature in school evaluation. Standards suggested for judging school evaluation included those addressing accuracy, utility, propriety, and feasibility of the evaluation.

A. UTILITY STANDARDS

3. Information Scope and Selection. Information collected should be of such scope and selected in such ways as to address pertinent questions about the object of the evaluation and be responsive to the needs and interests of specified audiences.

Craig, Marilyn Martin. Assessing the Effectiveness of a Framework for the Identification of Information Needs in Program Evaluation. Paper presented at the annual meeting of the American Educational Research Association, 1979. 23p. ED 174 688.

The development and testing of the effectiveness of a model for the identification of information needs in program evaluation are discussed. More than 200 subcategories of information needs were divided into three major categories: history, conception, planning and development of the program; operation--the ongoing processes of the program; and impact--program results. Five major variables were investigated: (1) number of information needs identified; (2) type of information needs identified; (3) adaptability to varied evaluation situations; (4) subject ratings on importance and priority; and (5) comprehensiveness of the framework. Results verified the framework's potential as an effective tool to aid evaluators in exploring evaluation situations, in broadening the scope of evaluation studies, and in increasing the impact of those studies.

Hayman, John; And Others. On Aggregation, Generalization, and Utility in Educational Evaluation. Unpublished. 1979. 25p. ED 174 667.

The cross-levels hypothesis is presented as an explanation for program evaluation failures. It states that the usefulness of evaluation data as feedback for decision making varies inversely with the number of organizational levels between the action the data described, and the decisions they are intended to influence. To be useful for decision making, evaluation data must meet three hierarchical information needs: syntactic, semantic, and behavioral. Syntactic errors, evaluators should specify their level of reference--individuals, classes, districts, states, nations--and realize that aggregating data across levels may confuse relationships among variables. Evaluation data must be on the same level as decision-maker concerns, to satisfy semantic needs. Formal evaluation reports, for example, are not relevant to teachers. The behavioral need explains why decision makers are less motivated by evaluation data removed from their level--politically speaking, this information is not perceived as important to their own concerns. The cross-levels hypothesis is strongly supported by these information needs, and offers an alternative to research design or statistical procedures as an explanation for program.

Metfessel, Newton S.; Michael, William B. A Paradigm Involving Multiple Criterion Measures for the Evaluation of the Effectiveness of School Programs. Educational and Psychological Measurement, v27 n4 pt2 p931-943, Winter 1967.

The twofold purpose of this paper is (1) to present an eight-step procedural outline of the evaluation process and (2) to furnish a detailed listing of multiple criterion measures that may be used in the evaluation of specific behavioral objectives. The eight major steps in the evaluation process are: direct and indirect involvement of the total school community as facilitators of program evaluations; formation of a cohesive model of broad goals, and specific objectives; translation of specific objectives into a communicable form applicable to facilitating learning in the school environment; instrumentation necessary for furnishing measures allowing inferences about program effectiveness; periodic observations of behaviors; analysis of data given by status and change measures; interpretation of data relative to specific objectives and goals; and recommendations culminating in further implementation, modifications, and revisions of broad goals and specific objectives. The measures which can be used to collect data include measures of student, teacher and community behaviors, collected through standardized tests, informal instruments, and other means.

Ott, Jack M.; And Others. Taxonomy of Administrative Information Needs: An Aid to Educational Planning and Evaluation. Educational Technology, v13 n5 p29-31, May 1973. EJ 079 052.

Since the quality of administrative decisions depends in part on the information the administrator has, incomplete or wrong information will be reflected in the decisions made. Thus, information needs must be anticipated in order that the gathering of that information may be planned. A nine-stage decision process is the basis for this taxonomy of information needs. At each stage of the decision process, the necessary information is specified. According to the taxonomy, the evaluation team is responsible primarily for locating present or potential inconsistencies and presenting them along with their probable causes and effects to administrators. It is assumed that the administrator will make the subdecisions that are involved in the decision process, such as establishing criteria for judging alternatives, or designing potential innovations. (The taxonomy is also available as ED 044 423).

A. UTILITY STANDARDS

4. Valuational Interpretation. The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

Anderson, Scarvia B.; Ball, Samuel. The Profession and Practice of Program Evaluation. San Francisco: Jossey-Bass, 1978. Pages 110-164.

This section on ethics and values in evaluation includes three chapters. The first argues that it is worthwhile for the evaluator to make explicit, in as honest and open a way as possible, his or her values. The second notes that the political-economic context of an evaluation also introduces bias, and this context must be taken into consideration as the evaluator forms relationships within and outside of the program being evaluated. The final chapter explores the ethnical responsibilities of the various players in an evaluation setting.

Apple, Michael W. The Process and Ideology of Valuing in Educational Settings. In Apple, Michael W.; And Others (Eds.), Educational Evaluation: Analysis and Responsibility. Berkeley, Calif.: McCutchan, 1974. Chapter 1.

Evaluation is a process of social valuing: it involves one or more groups of people assigning values to activities, goals, and procedures done by others, such as students. All too often, an evaluation is used to legitimate an educator's own common sense activity rather than to challenge it. Evaluation expertise often serves as an administrative procedure that is relatively ineffective in bringing about significant changes in educational processes. Because the choice of what one is to assess is itself a valuative decision, institutional evaluation (assessment of the quality of life students experience in schools) is often ignored.

Berk, Ronald A.; Rossi, Peter H. Doing Good or Worse: Evaluation Research Politically Re-Examined. Social Problems, v23 n3 p337-349, February 1976. EJ 142 091.

This paper argues that all evaluation research must necessarily rest on significant moral and political value judgments. These and other methodological factors in turn affect social problem definitional processes surrounding ongoing social programs. Moreover, evaluation research implicitly endorses particular ideological perspectives and therefore has broader implications for social change. Despite these serious weaknesses, however, evaluation research may play a progressive role if one is prepared to employ research designs that capitalize on inevitable value judgments, rather than ignore them.

Berlak, Harold. Values, Goals, Public Policy and Educational Evaluation. Review of Educational Research, v40 n2 p261-278, 1970.

An evaluation may focus on programatic or on public policy questions. Four criteria may be used to identify public policy issues: (1) Does the program alter the power relationship between the citizen and the state? Does it affect a person's status or power within the social system? (3) Does it increase or decrease social or political tensions? or (4) Does it effect a change in the self-concept of the individual? Public policy and programatic outcomes may be intended, unintended and anticipated, or unintended and unanticipated. This diversity of outcomes raises the boundary problem, i.e., the evaluator must determine which outcomes to study. The expert must set boundaries for a given evaluation task, and the determination of whether he or she will describe, recommend judgment criteria, or render a judgment depends upon whether the issue is primarily a public policy issue or a programatic issue.

Eichelberger, R. Tony. Multiple Stakeholders and Evaluation. Paper presented at the annual meeting of the American Educational Research Association, 1978. 20p. ED 164 565.

Evaluations occur within a political decision-making milieu, where multiple stakeholders are contending for limited funds. Given the subjective basis of empirical information, different conclusions or recommendations about a program may result from different ideological, theoretical, and disciplinary perspectives. The logic behind the interpretation of results, and the assumptions necessary for such interpretations, must be specified and explained to facilitate the most appropriate use of an evaluation. Because of the complexity of many statistical techniques presently used, much work is needed to identify what assumptions must be met for meaningful and useful interpretations of results in a specific decision-making situation. The rationales for both the inclusion and the exclusion of the variables to be considered in an evaluation should be made explicit. The problem of obtaining a matched control group is often nearly impossible. The relationship between the statistical analysis and the evaluation question is often based on tenuous assumptions. The evaluation of Project Follow Through is used to exemplify these problems.

Gorry, G. Anthony; Goodrich, Thelma Jean. On the Role of Values in Program Evaluation. Evaluation Quarterly, v2 n4 p561-572, November 1978. EJ 193 492.

When participants with varied background and interests join in a collaborative activity, their different viewpoints may make the evaluation of the activity more difficult. The emphasis placed on different kinds of success may differ greatly among project participants, causing them to disagree over the worth of the components of the program, irrespective of the technical merits of the evaluation of these components. An experience evaluating a multidisciplinary biomedical research center illustrates the influence of values on program evaluation.

Gross, Alan L. Funding Education Projects: Applying Decision Theory to the Problem. In Abramson, T.A.; And Others (Eds.), Handbook of Vocational Education Evaluation. Beverly Hills, Calif.: SAGE Pub., 1979.

Two decision theories (Multiattribute Utility Theory and Bayesian Decision Theory) that have been employed by educational evaluators to assist decision makers in making funding decisions are described. Both approaches are illustrated in terms of hypothetical, although realistic, examples.

Guttentag, Marcia. Subjectivity and Its Use in Evaluation Research. Evaluation: A Forum for Human Service Decision-Makers, v1 n2 p60-65, 1973.

Edwards' multi-attribute utilities model for evaluation is described, and its application in an Office of Child Development evaluation is discussed. The model, which quantifies decision-makers' values, is based on decision theory, and on the belief that the inherent subjectivity of decision-making makes classical experimental designs not very useful.

House, Ernest. Context and Justification. In Hamilton, David; And Others (Eds.), Beyond the Numbers Game: A Reader in Educational Evaluation. London: Macmillan Education, 1977. Section 4, Chapter 9.

A distinction is made between the "context of valuation" and the "context of justification." The context of valuation involves the basic value slant derived from the genesis of the evaluation, and includes all those motivations, biases, values, attitudes and pressures from which the evaluation arose. The context of justification involves the attempt to justify the findings. Utilizing scientific methodology in the context of justification enables one to minimize bias, but not eliminate it. Since all biases cannot be eliminated, it is essential that the scientist reveal the values on which his or her research is based.

House, Ernest R. The Conscience of Educational Evaluation. Teachers College Record, v73 n3 p405-414, February 1972. (Also reprinted in House, Ernest R. (Ed.), School Evaluation: The Politics and Process. Berkeley: McCutchan, 1973. Chapter 10.)

A variety of evaluation problems are discussed. First, there is no real demand among teachers and administrators for evaluating their own programs, unless the evaluation has some direct value for them. In this context, evaluations can be used for defense of a program or for attack of another program. It is useful to distinguish between the context of valuation (the value slant derived from the genesis of the evaluation) and the context of justification (an attempt to justify the findings). Using scientific methodology and making valuations explicit will allow evaluations to be as unbiased as possible. A final check on the evaluator's valuations and biases is the interests of other people.

House, Ernest R. Justice in Evaluation. In Glass, Gene V. (Ed.), Evaluation Studies Review Annual. Volume 1. Beverly Hills, Calif.: SAGE Pub., 1976. Chapter 3.

The prevalent conception of justice in evaluation is based on utilitarian ethics, i.e. the best endeavor is that which produces the greatest good for the greatest number. Rawls' conception of "justice-as-fairness" is proposed as an alternative. In this paradigm, each person is presumed to have nonnegotiable rights which cannot be bargained away no matter how it affects the good. The justice of several specific evaluation schema is reviewed.

Johnson, Mauritz. The Locus of Value Judgments in Educational Program Evaluation. Studies in Educational Evaluation, v5 n2 p109-122, 1979. EJ 211 901.

The purpose of this article is to examine the various decisions associated with evaluation, provide some sort of structure and terminology for them, and then determine where (within this structure) the crucial value judgments lie. Four types of decisions can be distinguished, based on whether they are internal or external, and instrumental or consummatory: authorizing decisions, consequential decisions, procedural decisions, and evaluation decisions. Among the various procedural decisions that must be made in designing and executing an evaluation are four that require value judgments: determination of criteria, criterial weights, standards, and rules for applying standards. A comprehensive definition of evaluation explicitly incorporates fact, value, and purpose: evaluation is (1) a judgment of the inherent or instrumental worth, (2) of some educational entity or process (evaluand), (3) for the purpose of enlightening an anticipated decision-making process, (4) arrived at by establishing explicit absolute or relative standards, (5) pertaining to relevant criteria or attributes of the evaluand, (6) that have been weighted in accordance with their perceived contribution to the evaluand's overall worth, (7) and applying the standards, according to appropriate rules, (8) to a full and accurage description of the evaluand, and (9) based on reliable observation pertinent to the criteria.

Krathwohl, David R. The Evaluator as Negotiations Facilitator-Fact Finder. Educational Evaluation and Policy Analysis, v2 n2 p25-34, March-April 1980. EJ 229 182.

Values are involved in every evaluation; the problem is to ensure that they are beneficial values, beneficially applied, and so perceived by the sponsor and relevant audiences. The fact that what is viewed as beneficial by one person or group may not be so viewed by another, makes clear the difficulty of trying to get evaluations accepted and used when we concentrate solely on the technical aspects. When audiences disagree on what is beneficial, there must be negotiation to reach an agreement on how the evaluation can be made most mutually beneficial. It is only as this aspect of evaluation is understood and resolved by the parties interested in it that the evaluation will be perceived as acceptable and extensively used by them in the decision-making process.

Krathwohl, David R. The Myth of Value-Free Evaluation. Educational Evaluation and Policy Analysis, v2 n1 p37-46, January-February 1980.

Values are and must be involved in evaluation. The choice of evaluation as a useful process, the definition of its role, what is studied, how resources are allocated, all involve value judgments. The problem is one of determining what is "beneficial prejudice" in any given instance.

Kunkel, Richard C.; Tucker, Susan A. A Perception-Based Model of Program Evaluation: A Values Oriented Theory. Paper presented at the annual meeting of the American Educational Research Association, 1977. 23p. ED 152 809.

Personnel at Saint Louis University's Department of Education have been theorizing, researching, and enlarging an approach to program evaluation that focuses heavily on the place of values in making judgments. This work originally stemmed from general curriculum evaluative theory developed by James H. McElhinney and Richard C. Kunkel. The content of the theory presented here contains: arguments for a theory of evaluation with explicit quality criteria; some quality criteria currently part of the perception-based model; a statement of the theory and operational paradigm; some polemics developing as the theory and operational paradigm; some polemics developing as the theory is being applied; and a brief section summarizing the theory's applications. In the perception-based model of evaluation proposed, certain value criteria are not open to negotiation in the sense that along with accepting the evaluator personally, the primary audience must accept five quality criteria inherent in the model: holism, helpfulness toward program improvement, evaluator vulnerability, acceptance of both "hard (observable) and "soft" (subjective) data sources, and facilitation of planning a program's future. Additional quality criteria are negotiated with individual audiences.

Lincoln, Yvonna S.; Guba, Egon G. The Distinction Between Merit and Worth in Evaluation. Paper presented at the annual meeting of the Evaluation Network, 1979. 21p. ED 183 574.

Valuing in evaluation encompasses two distinct senses of the word, denoted by the terms merit and worth. Merit may be defined as an entity's inherent, intrinsic, context-free value, while an entity's worth is defined as its contextually determined, place-bound value. Determining an entity's merit may take place whenever a number of experts are assembled. Worth can only be determined by viewing the entity in operation or on site. Thus, while merit may be determined in any number of ways, worth can be determined only by intensive field studies on site. And field studies often call for naturalistic, not scientific, approaches. Although it would seem that merit and worth are identical to formative and summative dimensions, they are orthogonal. It is therefore possible to create a 2 X 2 table and generate four distinct types of evaluation: formative merit evaluation, formative worth evaluation, summative merit evaluation, and summative worth evaluation. Each of the four types of evaluation serves distinctly different purposes and is addressed to different audiences and stakeholders.

Messick, Samuel. The Standard Problem: Meaning and Values in Measurement and Evaluation. American Psychologist, v30 n10 p955-966, October 1975. EJ 125 292.

The term "standard" in the title of this article is intended not only in its common dictionary meaning of "something established for use as a rule or basis of comparison in measuring or judging capacity, quantity, content, extent, value, quality, etc.," but also in its more general dictionary meaning of "something used by general agreement to determine whether or not a thing is as it should be." Accordingly, this article deals not only with questions of meaning but also with questions of values in both measurement and evaluation.

Myrdal, Gunnar. Objectivity in Social Research: The 1967 Wimmer Lecture, St. Vincent College, Latrobe, Pennsylvania. New York: Pantheon Books, 1969.

The most fundamental methodological problems facing the social scientist are: What is objectivity, and how can the researcher attain objectivity in trying to find out the facts and the causal relationships between facts? The logical means available for protecting ourselves from biases are broadly these: to raise the valuations actually determining our theoretical as well as our practical research to full awareness; to scrutinize them from the point of view of relevance, significance, and feasibility in the society under study; to transform them into specific value premises for research, and to determine approach and define concepts in terms of a set of value premises which have been explicitly stated.

Scriven, Michael. The Concept of Evaluation. In Apple, Michael W.; And Others (Eds.), Educational Evaluation: Analysis and Responsibility. Berkeley, Calif.: McCutchan, 1974. Chapter 3.

Evaluation involves making value judgments. The ultimate problem about evaluation is where the values come from. Value is a complicated theoretical turn that implies and follows from various combinations of desires, needs, and performance. Evaluation should be thought of as a process of compressing complex data in the vise of these contextual constraints so as to squeeze out the water and leave behind the meaty residue of directed information that is a value judgment. These value judgments are nearly always implicitly comparative if not explicitly comparative, and a clear recognition of this leads to important practical improvements in the utility of value judgments.

Sjoberg, Gideon. Politics, Ethics and Evaluation Research. In Guttentag, Marcia; Struening, Elmer L. (Eds.), Handbook of Evaluation Research. Volume 2. Beverly Hills: SAGE Pub., 1975. Chapter 3.

Research design should be reconceptualized to take account of social factors that structure the research from its inception on through the analysis of the findings. The relationship between the researcher as a variable in the research design and theory building is most clearly seen when the impact of the researcher's assumptions about human nature and social reality upon the research process is recognized. Although content, they also have a responsibility to science and to the principle of human dignity to recognize the broader political and ethical implications of their efforts.

Smith, Nick L. Sources of Values Influencing Educational Evaluation. Research, Evaluation, Development Paper Series No. 7. Portland, Oreg.: Office of Research and Evaluation Services, Northwest Regional Educational Lab., May 1977. 39p. ED 161 889.

With the theory that social and personal values influence the conduct of evaluation studies in education, the author discusses the impact of two major sources of such values: contextual factors, including political, social and organizational influences; and the terminology, models, and personal values of evaluators. Alternative purposes for an evaluation study are discussed and illustrated. In addition, values hidden in terminology, value-laden evaluation models, an evaluators' personal values and an illustration of evaluator roles are treated. The benefits and problems of values are discussed and four means of clarifying values in evaluation work are outlined. The first approach suggests that all relevant value positions need to be identified and stated publicly. The second approach emphasized the need to clarify the evaluator's role in the assessment process. Is he/she describing the program, recommending evaluation criteria, or rendering an actual judgment of worth? Through identifying his/her role, the evaluator can choose to de-emphasize his or her personal values. The third approach suggests explicitly incorporating opposed values into evaluation studies by conducting comparative analysis. The fourth approach reflects attempts to search out conflicting value positions to insure an appreciation of the full range of potentially influential values.

A. UTILITY STANDARDS

5. Report Clarity. The evaluation report should describe the object being evaluated and its context, and the purposes, procedures, and findings of the evaluation, so that the audiences will readily understand what was done, why it was done, what information was obtained, what conclusions were drawn, and what recommendations were made.

Brager, Gary L.; Mazza, Paul. The Level of Analysis and the Level of Presentation Are Not the Same. Educational Evaluation and Policy Analysis, v1 n3 p105-106, May-June 1979. EJ 211 827.

Suggestions are made on effective presentations by evaluators of research studies to audiences who are not statisticians. Examples of effective presentation methods are given: analogies in presenting statistics; graphs or pictorial presentations; summaries to highlight findings; concise reports based on a television newscast style; and judicial use of statistics.

Brown, Robert D.; And Others. Evaluator Credibility as a Function of Report Style: Do Jargon and Data Make a Difference? Evaluation Quarterly, v2 n2 p331-341, May 1978. (A revised version is available as ED 137 388.)

The impact of professional jargon and data-based statements in evaluation reports on audience responses to an evaluation report and an evaluator's recommendations were examined. Subjects read one of four evaluation reports about testing and grading procedures in a school program. The reports varied in the amount of jargon and data used to justify the recommendations. Ninety-five high school teachers and administrators read one of four short reports, each containing one of the following types of statements: (1) jargon-loaded, objective; (2) jargon-free, objective; (3) jargon-loaded, subjective; and (4) jargon-free, subjective. The jargon-loaded reports were rated as more technical than the jargon-free reports. The least difficult format was the jargon-free subjective report and the most difficult was the jargon-loaded subjective report. The subjective reports were rated as more practical and the jargon-loaded subjective reports were rated as less believeable than jargon-free objective reports. There were no differences in reactions to the recommendations of the evaluator. The results suggest that the impact of an evaluation report depends upon the style in which it is written.

House, Ernest R. Coherence and Credibility: The Aesthetics of Evaluation. Educational Evaluation and Policy Analysis, v1 n5 p5-17, September-October 1979. EJ 215 210.

Evaluation studies are discussed in terms of aesthetic and literary qualities. Concepts such as imagery, coherence, credibility, dramatic structure, mode of presentation, and story line, are analyzed in relation to evaluate documents.

Popham, W. James. Reporting Evaluation Results. In Popham, W. James, Educational Evaluation. Englewood Cliffs, N.J.: Prentice-Hall, 1975. Chapter 12.

Considerable attention must be given to the procedures employed to report the results of an educational evaluation. Most often, the evaluator supplies evidence to be used by others who make decisions, and so should assume a responsive orientation to these decision makers. A variety of techniques can be used to prepare the final report; for example, preparation of a work evaluation report, use of a diversity of reporting mechanisms, use of differential depth within a written report, use of adversary reporting techniques, use of communications specialists, summarization of the results, or provision of a preview copy of the report to those individuals whose program is being evaluated.

Roberts, Sarah. Communicating Evaluation Results. Module 12. Palo Alto, Calif.: American Institutes for Research in the Behavioral Sciences, 1978. 87p. ED 181 345. (Paper copy available only from National Consortium Project, American Institutes for Research, P.O. Box 1113, Palo Alto, CA 94302, $3.20.)

This module is the twelvth in a series on developing a comprehensive career guidance program at the high school level, designed to aid guidance personnel responsible for developing student-focused programs. The goal of this module is to help users develop the skills needed to produce an effective evaluation report in terms of content, format, level of sophistication, accuracy, and organizaiton. The module format consists of an overview, goals, objectives, outline, time schedule, glossary, readings, skill development activities, and bibliography. A coordinator's guide is also included with detailed instructions for presenting the module in a workshop setting, as well as the facilitator's roles and functions, and the criteria used in assessing the participants' achievement of module objectives.

Wolf, Richard M. Data Analysis and Reporting Considerations in Evaluation. In Popham, W. James (Ed.), Evaluation in Education: Current Applications. Berkeley, Calif.: McCutchan, 1974. Chapter 4.

The path from a collection of observations and measurements to a set of warranted conclusions is fraught with hazards. This chapter describes the path and offers some guidance on how to negotiate it. It also discusses presenting results in a way that can be understood by nontechnically trained persons. It should enable the reader to better identify and classify each variable in a study in terms of its status and scale of measurement; acquire information about the data to be analyzed; identify the stages of treatment of data; select an appropriate statistical procedure; and present the results of a statistical analysis in a way that can be understood by teachers, administrators, school board members, and parents.

A. UTILITY STANDARDS

6. Report Dissemination. Evaluation findings should be disseminated to clients and other right-to-know audiences, so that they can assess and use the findings.

Ball, Samuel; Anderson, Scarvia B. Dissemination, Communication, and Utilization. Education and Urban Society, v9 n4 p451-470, August 1977.

Dissemination involves more than just telling the world (or some subsection of it) what an evaluation has concluded. It should involve informing others about the evaluation plans, procedures, and later its findings. A number of different audiences should be included in the evaluation plan. The communications network for dissemination should include the evaluator, evaluation staff, program staff, and program participants. These groups should be in close communication throughout the evaluation. Finally, if an evaluation is to be utilized, evaluators must be advocates of their results, active in bringing them to the attention of others, and willing to identify publicly any policy and practical implications

Patton, Michael Quinn. The Meanings of Evaluation Data: Analysis, Interpretation, Dissemination, and Utilization. In Patton, Michael Quinn, Utilization-Focused Evaluation. Beverly Hills, Calif.: SAGE Pub., 1978. Chapter 11.

Evaluation research is ultimately a personal, perceptual, and interpretive approach to establishing the effectiveness of human service activities. To increase its utilization: (1) present the data in such a way that decision makers can decipher and interpret findings for themselves; (2) discuss and negotiate the format, style, and organization of final reports with those who will be the primary users of each report; (3) make dissemination efforts a matter for negotiation and cooperation between decision makers and evaluators as they work together to make study findings relevant and meaningful to various larger audiences; and (4) personalize evaluation reports by identifying both the evaluators who wrote the reports and the decision makers for whom they were written.

Stevens, W. F.; Tornatzky, L. G. The Dissemination of Evaluation: An Experiment. Evaluation Review, v4 n3 p339-54, June 1980.

The utilization of program evaluation methodology in human service agencies was reviewed from the perspective of organizational contingency theory. Adoption of program evaluation was seen as an innovation which would arouse uncertainty in an organization. A 2x2 factorial experiment, with a sample of 37 drug abuse programs, was conducted to test two hypotheses: (1) group consultations with staff would produce more innovation adoption than private consultations with a program director; and (2) on-site consultations with face-to-face interactions would produce more innovation adoption than telephone consultations. Results indicated strong support for the first hypothesis, and more ambiguous support for the second.

A. UTILITY STANDARDS

7. Report Timeliness. Release of reports should be timely, so that audiences can best use the reported information.

Anderson, Scarvia B. Dissemination of Evaluation Results. In Anderson, Scarvia B.; And Others, Encyclopedia of Educational Evaluation: Concepts and Techniques for Evaluating Education and Training Programs. San Francisco: Jossey-Bass, 1975. Pages 130-132.

Dissemination includes the issues of who should get the results, what kinds of results should be reported for what purposes, and when and in what form results should be reported.

A. UTILITY STANDARDS

8. Evaluation Impact. Evaluations should be planned and conducted in ways that encourage follow-through by members of the audiences.

Agarwala-Rogers, Rehka. Why Is Evaluation Research Not Utilized? In Guttentag, Marcia (Ed.), Evaluation Studies Review Annual. Volume 2. Beverly Hills, Calif.: SAGE Pub., 1977. Chapter 16.

The underlying reasons responsible for the underutilization of evaluation research are presented: lack of administrator involvement in the evaluation process, conflicting interests of program staff and evaluators, lack of mutually agreed upon "problem" definition, lack of special liaison staff between program staff and evaluators, lack of emphasis on providing solutions to problems, overemphasis on reporting negative findings, and problems of feedback and timeliness of results. Suggestions for increasing utilization of evaluation results include use of an evaluator who is an insider to the organization, involvement of program staff in the evaluation, and provision of liaison individuals or institutions to translate needs into evaluation research and evaluation research into practice.

Alexander, Jay; And Others. Increasing the Use of Evaluation Information: An Evaluator-Manager Interaction Model. San Antonio, Tex.: Education Service Center Region 20, February 1980. Paper presented at the annual meeting of the Southwest Educational Research Association, 1980. 14p. ED 185 040.

An evaluator-manager interaction model is presented for predicting the impact of evaluation and research findings. Instruments were developed for measuring the variables of interpersonal involvement, impact of evaluation, and managerial style in the relationship between evaluator and manager. The hypothesis advanced suggests that evaluators can improve their efficiency and impact of shifting the bulk of their interpersonal involvement towards managers who are more reluctant to use evaluation data to change their ongoing educational programs.

Brown, Robert D.; Braskamp, Larry A. Summary: Common Themes and a Checklist. New Directions for Program Evaluation, n5 p91-97, 1980. EJ 229 197.

Six common themes related to evaluation utilization as represented in the papers in this issue of New Directions for Program Evaluation, are summarized: the definition of utilization, the immediate concern of evaluators, the active role of evaluators in enhancing utilization, the relevance of evaluation information, relationship between evaluator and intended audiences, and the importance of the communication process. A 50-item Utilization Enhancement Checklist is presented, and covers five areas: determing evaluator role; understanding organizational context; planning; conducting the evaluation; and communicating evaluative information.

Ciarlo, James A. Utilizing Evaluation: Concepts and Measurement Techniques. SAGE Research Progress Series in Evaluation. Volume 6. Beverly Hills, Calif.: SAGE Pub., in cooperation with the Evaluation Research Society, 1981.

This collection focuses on utilization of the kind of information known as evaluation: whether program people absorb such information, and how, when, and what, if anything, they then do with respect to the programs they operate. Carol Weiss offers a taxonomy of methods to study different aspects of utilization and lists their strengths and weaknesses. John Stevenson describes an approach to assessing evaluation utilization in human service agencies. Ross Conner raises the issue of what organizational levels should be included among respondents being asked about utilization. Judith Larsen and Paul Werner discuss the utilization of consultants' suggestions for program improvement. Cathy Anderson, James Ciarlo, and Susan Brodie suggest the addition of effective utilization, or a change in emotional state or feeling about programs, to other types already identified by investigators. Finally, Donald Pelz and Jo Anne Horsley discuss the utilization of program-relevant research.

Cox, G. Managerial Style: Implications for the Utilization of Evaluation Information. Evaluation Quarterly, v1 n3 p499-508, 1977.

One of the central problems with program evaluation is the general perception that results are not utilized as fully as possible in decision-making processes. The fact that a similar problem exists in a wide range of information exchange situations suggests that the source of the problem is not primarily methodological. The article draws on Mintzberg's model of manager behaviors, and then draws some inferences as to how utilization would proceed and how it might be increased.

Davis, Howard R.; Salasin, Susan E. The Utilization of Evaluation. In Struening, Elmer L.; Guttentag, Marcia (Eds.), Handbook of Evaluation Research. Volume 1. Beverly Hills, Calif.: SAGE Pub., 1975. Chapter 20.

After reviewing a variety of research-development-dissemination utilization models, the authors present the human action model, or A VICTORY Technique. This paradigm is influenced by three concepts: the values of the individual, organization, or society; the capacity or ability to perform according to a selected idea; and prevailing circumstances and timing. The four steps in the use of the A VICTORY technique (assessment, goal definition, action, and follow-through) are described and discussed.

Granville, Arthur C.; And Others. The Impact of Evaluation: Lessons Drawn from the Evaluation of Five Early Childhood Education Programs. Paper presented at the annual meeting of the American Educational Research Association, 1978. 45p. ED 166 212.

Five different program evaluations are described to indicate those qualities which make an evaluation effective or not effective. Evaluation effectiveness was defined as impact on decision making or long-term policy formation, and influence upon a variety of audiences. Robert D. Matz described the First Chance Project, and concluded that the evaluation methodology used to inform policymakers should be distinct from the approach used to improve teaching. John M. Love, who was associated with the national Home Start Demonstration Program, felt that several factors contributed to the value and use of evaluation information: evaluations planned with the program; timely reports; rigorous experimental design; respect between agency and evaluators; demonstrated relationships between process and outcomes; and non-controversial appeal of the program. Project developmental continuity was discussed by Arthur C. Granville. Factors affecting evaluation impact included the relevance of quantitative data; sociopolitical acceptability of the implications; and pertinence to evaluation criteria. Allen G. Smith, who discussed Project Follow Through, supported close relationships between evaluation research and curriculum. Lawrence J. Schweinhart of the Ypsilanti (Michigan) Perry Preschool Project recommended intensive on-site studies, determination of the feasibility of longitudinal studies, and adequate funding.

Grobe, Robert P. Evaluation--What's It All About? NASSP Bulletin, v62 n422 p1-14, December 1978. EJ 192 364.

An increased emphasis has been placed on planned evaluation due to increased accountability needs, the large number of federally-funded projects, and the professional needs of educational administrators for better decision-making information. Five problems opposing the effective use of evaluation information include: (1) ambiguity of outcomes, (2) decision-makers who are unfamiliar with data, (3) emotional involvement with projects, (4) trivial evaluation requirements of federal government, and (5) expecting precise answers. The basic purpose of evaluations is to provide the administrator with an information base for decision making, including well-defined priorities, budgeting based on need, better planning, more efficient operations, more effective selection of special projects, and more state and federal funds. Stufflebeam's Context-Input-Process-Product (CIPP) model is used to demonstrate techniques to improve evaluation utilization.

Guba, Egon G. Problems in Utilizing the Results of Evaluation. Journal of Research and Development in Education, v8 n3 p42-54, Spring 1975. EJ 118 545.

This paper delineates some of the more frequently encountered utilization problems. First, conflict may arise between the overt foci of the evaluation and the covert foci, such as compliance or ratifying a decision already made. Second, the evaluation may not meet the criteria of a good evaluation: internal validity, external validity, reliability, objectivity, relevance, importance, scope, credibility, timeliness, pervasiveness, and efficiency. A third source of difficulty stems from discrepancies between program plans and actual operations. Innate differences among the many audiences entitled to receive the evaluation information may hinder utilization. A fifth source of difficulty stems from the rapid pace of change, both societal change and change in program mission. A sixth source of difficulty may result when the evaluator does not maintain a position of integrity with regard to the program. Finally, the sociopolitical context of the evaluation may hinder its use.

Haenn, Joseph F. Reasons Why Evaluations and Testing Don't Inform. Durham, N.C.: NTS Research Corp., April 1980. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980. 26p. ED 187 733.

A number of organizational, personal and methodological characteristics have been identified through the literature which inhibit or can be used to facilitate the use of evaluation and testing information. Inhibitors of information usage include organizational characteristics such as loosely coupled and decentralized systems, personal characteristics such as lack of awareness of interest in the needs and values of evaluation, and methodological characteristics of the evaluation. Although a few of these characteristics are static and not easily changed, most can be modified through the strategies of (1) creating a demand for the utilization of evaluation and testing information, (2) facilitating cooperation between evaluation personnel and decision-makers, and (3) improving reporting practices. A model of local district use of evaluation and testing information based on these characteristics and strategies is presented and discussed.

Mann, Floyd; Likert, Rensis. The Need for Research On the Communication of Research Results. In Caro, Francis G. (Ed.), Readings in Evaluation Research. Second Edition. New York: Russell Sage Foundation, 1977.

Based on data collected in the Detroit Edison Company in 1948, four factors were identified which are important for securing maximum acceptance and utilization of survey results: (1) a high degree of participation and personal involvement is important; (2) group forces are important in facilitating attitude changes and redefinitions of situations; (3) it is important to recognize the hierarchical structure of an organization; it is also essential to understand and utilize the power structure as perceived by the members of the organization; and (4) participation in a form of self-analysis is more likely to be followed by changes than if the analysis is made by an outsider.

Novak, Carl D. An Involvement Approach to the Evaluation of Local District Programs. Paper presented at the annual meeting of the American Educational Research Association, 1977. 46p. ED 152 819.

Ways to increase the use of educational program evaluation findings through the meaningful involvement of potential users (teachers and administrators) are discussed. Involvement, as defined in this paper, is generally limited to the opportunity for input to the initial evaluation design and the chance to review the design and implementation plans prior to the evaluation. Therefore, the role of staff, teachers, and administrators is to provide direction for the study. For effective, meaningful involvement, the following guidlines should be followed: (1) involve only individuals who can contribute something or have a stake in the program; (2) screen out biased inputs; (3) use the solicited information; (4) involve teachers and administrators in the program planning and implementation, but do not hold them responsible for the evaluation; (5) do not unnecessarily inconvenience or overburden the participants; (6) keep the evaluation planning process open; and (7) keep the audience informed of current progress. The evaluation of the Orton-Gillingham reading disabilities program used in Lincoln, Nebraska is described in detail.

Patton, Michael Q.; And Others. In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research. Minneapolis: Center for Social Research and Dept. of Sociology, Univ. of Minnesota, 1975. 46p. ED 135 938.

Research on the utilization of evaluations was based on a followup of 20 federal health program evaluations to assess the degree to which the evaluations had been used to to identify the factors that affected varying degrees of utilization. Interviews were conducted with project officers or people they identified as decision-makers who would utilize information in the evaluation reports. Two major themes emerged from the study. First, it was found that much of the evaluation literature has considerably overestimated the kind of impact evaluation research is likely to have. Second, the importance of the personal factor in evaluation research, particularly the utilization process, has been considerably underestimated. The two themes are directly linked. The impact of evaluation research is most often experienced as a reduction in the uncertainty faced by individual decision makers as they attempt to deal with the complexity of programing reality. It must be assimilated and fitted into a contextual whole. Energetic and interested people in government can and do use evaluation research, not for making decisions with immediate, concrete, and visible impacts, but in a more subtle, clarifying, reinforcing, and reorienting way. Evaluators, then, might do well to spend less time lamenting their lack of visible impact on major decisions and more of their time providing relevant information to those key persons whose thoughts and actions, to a substantial extent, determine the general direction in the evolutionary process of program development. It is in consciously working with such decision-makers to answer their questions that the utilization of evaluation research can be enhanced.

Tittle, Carol Kehr; And Others. A Procedure to Link Evaluation and Funding Decisions. Educational Evaluation and Policy Analysis, v3 n3 p43-53, May-June 1981.

Reported in this paper is the development and initial feasibility study of a set of procedures designed to establish a relationship between evaluation findings and funding decision-making. The decision-making setting was an annual grant program for vocational education administered at the state level. The procedure to link evaluation and funding decisions required determining the priority and criterion weights for major predictive and outcome impact variables, and providing estimates of the categories in which projects might be described for each of the impact scales. Feasibility was examined by surveying local education agencies to obtain data for each variable or to give an indication of future availability of data. This study demonstrated that evaluation findings and funding decisions can be linked to make better estimates of both predictive and outcome impact of projects.

Tittle, Carol Kehr. Evaluation and Decision Making: Developing a Method to Link Program Funding Decisions and Outcome Evaluation. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978. 14p. ED 155 219.

There is a continuing need in evaluation research for the establishment of a relationship between evaluation findings and decision making. A method is proposed for a particular situation: annual funding decisions for projects in a large grant program in vocational education. Outcome and predictive impact variables were ranked by three groups of decision makers on a pilot study. The groups included the Director of the State Department of Education division responsible for funding decisions, the supervisors who make funding decisions, and the supervisors from related bureaus who review and contribute to the decision-making process. Statements concerning the impact of vocational education programs on students, employers, and the State Department of Education--to be used as program evaluation criteria--were sorted into twelve outcome impact and nine predictive impact statements. Each statement was ranked and rated for importance by the decision-makers. Results showed high agreement on the ranking and rating of outcome impact statements, and discrepancies on the predictive impact statements. A validation study has been designed. Evaluators can assist decision-makers in identifying important outcomes; and in the process, define the decision to be made, the time when it is made and the data required to link evaluation and decision making.

Weiss, Carol H. Utilization of Evaluation Results. In Weiss, Carol H., Evaluation Research: Methods for Assessing Program Effectiveness. Englewood Cliffs, N.J.: Prentice Hall, 1972. Chapter 6.

Five constraints which frequently limit the use of evaluation results are discussed: (1) the evaluator's perception of her or his role in the evaluation process; (2) the organization's resistance to change; (3) inadequate dissemination of results; (4) the gap between evaluation findings and clear courses for future action; and (5) the tendency of much evaluation to show little or no positive effect. In each case, approaches for improving utilization are discussed.

B. FEASIBILITY STANDARDS

1. Practical Procedures. The evaluation procedures should be practical, so that disruption is kept to a minimum, and that needed information can be obtained.

Boruch, Robert F. On Common Contentions About Randomized Field Experiments. In Glass, Gene V. (Ed.), Evaluation Studies Review Annual. Volume 1. Beverly Hills, Calif.: SAGE Pub., 1976. Pages 158-194.

The resistance to a randomized comparative experimental design to answer the impact or effectiveness question about a program is deep-set and vigorously rationalized. Critics hold that experiments are impossible to implement in the "real world"; they are expensive and slow, they can be replaced by merely statistical adjustment of nonexperimental data, they are unethical, or that they ignore individual variance and idiosyncracy. In this paper, Boruch has martialled the rebuttals to these criticisms and has avoided reconstructing the opposing views as easily slain straw men.

Casper, Paul N.; Roecks, Alan L. Practical Program Evaluation. Paper presented at the annual meeting of the Southwest Educational Research Association, 1980. 36p. ED 182 303.

The practical side of a program evaluation, as performed at a Texas Education Service Center, is described. The role of the evaluators, as perceived by the users of the evaluation, and the procedures for evaluating programs several levels away from students who are to feel the effects of the evaluation, are discussed.

Powers, Donald E.; Alderman, Donald L. Practical Techniques for Implementing True Experimental Designs. Evaluation Quarterly, v3 n1 p89-96, February 1979. EJ 200 578.

It is sometimes possible to apply true experimental designs in field settings by taking advantage of the constraints under which programs or experimental treatments must operate. In a research study requiring classical treatment and control groups, practical methods for implementing true experimental procedures in public schools had to be devised and applied. These solutions to a problem often encountered by evaluators are presented here.

Wick, John W. On Evaluating a Project: Some Practical Suggestions. NCME Measurement in Education, v6 n1 p1-8, Winter 1975. (Also available as ED 109 167).

Prime indicators for realistic short term/long term project goals are budgets and timetables. Concrete, identifiable objects are useful in separating eloquent rhetoric from actual promises. Similarly, an external evaluator should be able to separate proposals with intentional misrepresentation of funding and goals from those which need further organization. Once a project begins, the evaluator should know whether the data being collected and analyzed will be used for internal public consumption, external public relations, or both. This may depend on whether the evaluators' primary allegiance is to the funding agency or to the project. In any evaluation, traditional staff roles and lines of authority should be recognized and better communication facilitated. Technical expertise and the political realities of a system should be reconciled.

B. FEASIBILITY STANDARDS

2. Political Viability. The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.

Banner, David K.; And Others. The Politics of Evaluation Research. In Banner, David K.; And Others, The Politics of Social Program Evaluation. Cambridge, Mass.: Ballinger Pub., 1975. Chapter 3.

Evaluation has explicit political overtones. It is designed to yield conclusions about the worth of a given social action program, and in so doing, it is intended to affect the allocation of resources. The purpose of this chapter is to examine the major literature in the politics of evaluation with an eye toward building a model of the process involved. This "model" explores the dimension of political interaction in the evaluation of social action programs.

Brickell, Henry M. The Influence of External Political Factors on the Role and Methodology of Evaluation. In Cook, Thomas D.; And Others (Eds.), Evaluation Studies Review Annual. Volume 3. Beverly Hills, Calif.: SAGE Pub., 1978. Chapter 5.

Several examples of external political influences on actual evaluations are presented. Five guidelines can be used to cope with such influences: (1) try to understand how the client thinks; (2) reassure the client that you can interpret the findings so as to give helpful suggestions for program improvement; (3) find out what the decision-makers will actually use as criteria for judging the success of the project; (4) try to get a supervisory mechanism set up for the evaluation contract that contains a cross-section of all the powerful decision makers; and (5) write the report carefully, especially when describing shortcomings or placing blame.

Caro, Francis G. Issues in the Evaluation of Social Programs. Review of Educational Research, v41 n2 p87-114, April 1971. EJ 038 655.

This paper reviews the literature on the use of the concepts and methods of behavioral research in evaluating social programs. The first part of the paper is concerned with basic issues which include definitions, approaches to evaluation methodology, roles of evaluation in program development, and distinctions among various forms of research. The second section deals with organizational matters such as the establishment of the evaluative research role, administration of evaluative research, utilization of the results of evaluation, and implications of client activism for evaluation. This section also includes a discussion of the basic tensions between evaluative researchers and administrators: service vs. research, specificity vs. generality, methods, status quo vs. change, explanations of failure, and academic vs. practical experience. The third section reviews methodological issues in measurement and design of evaluation studies.

Englert, Richard M.; And Others. Politics of Program Evaluation in Large City School Districts. Education and Urban Society, v9 n4 p429-450, August 1977. EJ 166 999.

This article explores some general notions about politics, evaluation, large city districts, and their interrelations. Politics, defined in terms of power, influence, policy conflict, and similar concepts, permeate every stage of program evaluation. Political forces are influential enough to give rise to the evaluation effort and to affect its implementation. At the same time program evaluation has an impact on political activities, especially policy-making. At times program evaluators themselves engage in political activities. These political activities are not necessarily unethical or inappropriate, but their existence should be recognized.

House, Ernest R. The Politics of Evaluation in Higher Education. Journal of Higher Education, v45 n8 p618-627, November 1974. EJ 107 533.

Analyzed are some of the political problems encountered in conducting evaluations in higher education. Liberal arts colleges have their own difficulties in which evaluation becomes entangled with suvival of the organization. Universities have difficulty using evaluation results because of the diffuse nature of their decision-making. Underlying many problems is the fact that projects are used to promote careers, and even moderate public statements about them can blemish the personal credentials necessary to career advancement. Within these constraints, the evaluator is necessarily in conflict to the degree he or she discovers flaws. Finally, an operational university evaluation system which minimizes some of these problems is cited.

Levine, Adeline; Levine, Murray. The Social Context of Evaluative Research: A Case Study. Evaluation Quarterly, v1 n4 p515-542, November 1977.

Evaluation takes place in a social context that influences research design, selection of variables, the written report, and the timing of its release. There are also consequences for program implementors, for those subject to the program, and for evaluators. Evaluations and evaluators may become involved in political conflict within the subject system and conflict external to it as well. The present study makes use of archival data to illustrate the issues in evaluations of the Gary plan of education that took place between 1914 and 1918. Suggestions for confronting political and social realities surrounding evaluation emerge from an application of concepts deriving from the sociology of knowledge.

Mathis, William. Evaluating: The Policy Implications. Paper presented at the annual meeting of the American Educational Research Association, 1980. 14p. ED 189 123.

Whether initiated by law, regulation, or administrative direction, evaluation has political purposes. Improvement, the classic purpose of evaluation, is most faithfully observed when the importance, funding, and constituency of a program are small. If something is wrong, or if the program's existence is threatened, evaluation can become a weapon in policy disputes, hiding values from constituents in a mystique of scientific inquiry. Purposes may also be reflected in the biases of those who initiate and conduct evaluations, in the selection of a program and objectives of evaluation, and in the amount of evaluation funding relative to program funding. Similarly, problems inherent to evaluation methods are sources of bias. Traditional quantitative measures do not lend themselves to broad and sweeping social programs, such as bilingual education. Finally, evaluation results are often used selectively to further political ends. In conclusion, evaluation can be viewed as an historical enterprise which seeks to recreate the past with selected emphases or biases.

Polemeni, Anthony J. The Politics of Evaluation. Paper presented at the annual meeting of the American Educational Research Association, 1978. 11p. ED 163 041.

At one time the major problems faced by an evaluator involved the best way to collect, analyze, and report data. Today, an evaluator's major problems concern responses to the evaluation report by school superintendents, principals, teachers, unions, and parents' groups. An unwillingness to publicize the evaluation results; the failure to consider evaluation results when making program decisions; the suppression of evaluation data if they adversely affect patronage possibilities; the demand for gross over-simplification in reporting evaluation results; an inability or reticence to create new programs, or to alter old ones based upon evaluative data; and the lack of communication between evaluators and field personnel are discussed as problems resulting from the political forces which influence evaluation. The role of laymen, supervisory personnel, project managers, classroom teachers, unions, the media, and parents in the politics of evaluation is outlined. Politically motivated critics of evaluation ignore the mandated necessity of assessment, and sometimes expect that evaluation agencies will be able to provide immediate data upon request. The author maintains that these political factors result in a lack of funding and facilities necessary in the work of a competent evaluator.

Sroufe, Gerald E. Evaluation and Politics. In Scribner, Jay D. (Ed.), The Politics of Education: The Seventy-Sixth Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1977.

Politics and evaluation are intimately related. Politics has to do with the distribution of stakes within a society or group; evaluation is oriented toward improved decision-making, and its goal is a judgement of value, worth, or merit. Evaluation is a political resource that can be used to influence the distribution of stakes in education. Evaluation can be offensive (undertaken to alter the existing distribution of stakes) or defensive (designed to thwart an offensive evaluation). Finally, in any study, the politics of the individual and the evaluation agency must both be considered.

Tumin, Melvin M. Politics of Evaluation. In Anderson, Scarvia B.; And Others, Encyclopedia of Educational Evaluation: Concepts and Techniques for Evaluating Education and Training Programs. San Francisco: Jossey-Bass, 1975. Pages 281-286.

Political conflicts in program evaluation can arise over which goals are most and which least important, which values are to be preserved and which can be sacrificed, what is the acceptable ratio of cost and effort to gain and achievement, what will determine whether the program has succeeded or failed, and who shall make such judgments. The politics of evaluation refers to any partisan activities directed at influencing the conduct of evaluation in line with partisan preferences. So understood, political considerations may and almost always do enter into evaluation at every stage.

Weiss, Carol H. Evaluation Research in the Political Context. In Struening, Elmer L.; Guttentag, Marcia (Eds.), Handbook of Evaluation Research. Volume 1. Beverly Hills: SAGE Pub., 1975. Chapter 2.

Evaluation is a rational enterprise that takes place in a political context. Political considerations intrude in three ways: (1) the policies and programs with which evaluation deals are the creatures of political decisions; (2) because evaluation is undertaken in order to feed into decision-making, its reports enter the political arena; and (3) evaluation, by its very nature, makes implicit political statements.

Wright, William J. Comments on "The Influence of External Political Factors on the Role and Methodology of Evaluation." In Cook, Thomas D.; And Others (Eds.), Evaluation Studies Review Annual. Volume 3. Beverly Hills, Calif.: SAGE Pub., 1978. Chapter 6.

The primary issue raised by this paper is the need to examine ways of solving the problems resulting from the inevitable intertwinement of politics and evaluation. Standards for evaluation should be generated that are consistent with the commonly-held values of evaluators. There are two general areas in which standards might be generated: contracts (respective responsibilities, audience restrictions, conflict of interest), and performance (instrumentation and sampling, interpretation and reporting, nonperformance).

B. FEASIBILITY STANDARDS

3. Cost Effectiveness. The evaluation should produce information of sufficient value to justify the resources expended.

Cost of Educational Accountability--A Maryland Exploratory Study. Denver: Cooperative Accountability Project, Colorado State Dept. of Education; Baltimore: Maryland State Dept. of Education, 1974. 62p. ED 102 722.

The Maryland State Department of Education participated with the Cooperative Accountability Project (CAP) in an exploratory study of the cost-pricing of educational accountability components. The exploratory study was undertaken to determine the state of the art in cost-pricing of accountability components at the state and local educational levels and to enable the organizations to make recommendations about necessary, future research in this field. Four educational accountability components were identified: goal development and implementation, objective development and implementation, status surveying of student achievement, and program development. Based on these components, a survey instrument was constructed to obtain information from local school systems about the costs involved in actually providing information to decision-makers. The basic conclusion reached in the survey is that smaller school systems will require additional financial aid and technical assistance in establishing a comprehensive accountability program.

Schriber, Peter E. Cost Benefit Analysis of Comprehensive Achievement Monitoring for Classroom Evaluation. Amherst: School of Education, Univ. of Massachusetts, February 1971. Paper presented at the annual meeting of the National Council on Measurement in Education, 1971. 11p. ED 053 181.

Comprehensive Achievement Monitoring (CAM) is a systematic procedure of constructing and administering longitudinal, criterion-referenced tests. CAM has advantages over typical classroom testing of having well-organized tests, providing course evaluation through pretesting, posttesting, and retention measurement; producing data for continuous classroom instructional management; and being based on a curriculum of behavioral objectives. The systematic gathering of comprehensive performance data permits a means of establishing a dollar-and-cents cost analysis for various instructional and curricular alternatives. The costs of typical classroom testing and CAM are compared.

C. PROPRIETY STANDARDS

1. Formal Obligation. Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

House, Ernest; And Others. An Assessment of the Michigan Accountability System. March 1974. 64p. ED 099 21.

Michigan assumed a leadership role in exploring and applying accountability procedures. The purpose of this report is to examine the quality and implications of that leadership. Specifically, it assesses the Michigan Accountability System with respect to its educational soundness and utility for Michigan, and with particular emphasis on the assessment component. The report presents both positive and negative findings organized by the criteria used to assess an accountability program. It also includes a copy of the memorandum of agreement between the evaluators and the sponsors (the Michigan Education Association and the National Education Association). A staff response to this report is available as ED 111 838; and Stufflebeam's response to the staff response is available as ED 163 058.

A Staff Response to the Report: An Assessment of the Michigan Accountability System. Lansing: Michigan State Dept. of Education, May 1974. 40p. ED 111 838.

This response was made to an evaluation of the Michigan Accountability System (House, ED 091 821). Ernest House, Wendell Rivers, and Daniel Stufflebeam were contacted by the Michigan Education Association and the National Education Association to evaluate the System's educational soundness and utility, with a particular focus on the assessment component. To some extent, the study produced observations and judgments without inaccuracies or emotional exhortations. However, the original report contained some inaccuracies, it was not totally unbiased, and it appeared to be based on somewhat unrigorous and hurriedly-gathered information. Problem areas included observations on goals and objectives, state level leadership, testing, teacher evaluation, and the compensatory education program. Stufflebeam's response to the staff response is available as ED 163 058.

Stufflebeam, Daniel L. A Response to the Michigan Education Department's Defense of Their Accountability System. Paper #1 in Occasional Paper Series. Kalamazoo: School of Education, Western Michigan Univ., August 1974. 36p. ED 163 058. (Hard copy available only from the Evaluation Center, College of Education, Western Michigan University, Kalamazoo, MI 49008.)

The author responds to reactions by program personnel to an evaluation of their program conducted by him and others. The program was the Michigan Accountability System. The program was conducted by the Michigan Department of Education and it was evaluated by the author and others who were contracted for the work by the Michigan Education Association and the National Education Association.

The author reviews the history of his agreeing to do the evaluation, reviews the evaluation findings, presents the program participants' reactions to the findings, and responds to their reactions. A written set of working agreements used to govern the study that was agreed to by all parties involved with the evaluation (evaluators, sponsor, program personnel) prior to the initiation of the work is included.

Weiner, Stephen S.; Rose-Pendleton, M. K. Separate Realities: A Case Study of Disagreement in the Design of an Evaluation. First Task Final Report. June 15, 1977. 74p. ED 152 814.

The National Institute of Education (NIE) had commissioned an evaluation project of certain postsecondary programs for nontraditional students that would involve decision-makers in the core of the design activity. This report discusses the conflict that emerged between NIE and the Center for Research and Development in Higher Education (CRDHE) and the failure of their efforts to resolve the conflict. The terms of the initial agreement did not establish a clear priority between the design of an evaluation and the necessity of consultation with a decision-maker, nor were the steps spelled out in operational terms. The Center's intellectual interests and their ties inclined them to questions of direct interest to program managers. NIE, however, was primarily interested in serving federal decision-makers' assessment as to how well specific groups were being served. The existence of this fundamental divergence did not become evident until it was too late to change the project. Furthermore, there was a clash of styles. CRDHE was accumstomed to a collegiate style and relied little on hierarchical lines of authority.

Wright, William J.; Worthen, Blaine R. Standards and Procedures for Development and Implementation of an Evaluation Contract. Portland, Oreg.: Northwest Regional Educational Lab., October 1976. 126p. ED 127 341.

Intended for individuals and/or agencies who provide or require evaluation services, this paper attempts to deal with the lack of standards and procedures for evaluation contracts. The first section of the report, Summary and Overview of Standards and Procedures for Evaluation Contracting, contains a brief discussion of the use of the proposed standards and procedures, a checklist proposed for use in applying the standards and procedures, instructions for use of the checklist, and a flowchart which shows the interrelationships and sequence of major events for applying the standards and procedures. The second section of the report, Rationale and Discussion Relevant to the Development of Standards and Procedures for Evaluation Contracting, extends the discussion and rationale referred to in the first section and is subdivided into the following subsections: (1) use of educational evaluation; (2) conceptual issues in determining when evaluation is appropriate; (3) rationale for use of external evaluation contracts; (4) rationale for specifying contractual procedures; (5) standards and procedures for selecting evaluation contractors; (6) standards and procedures for negotiation with an evaluation contractor; (7) standards and procedures for monitoring an evaluation contract; and (8) applications of the standards and procedures to sample contracts.

Wright, William J.; Worthen, Blaine R. Summary and Overview of Standards and Procedures for Evaluation Contracting. Portland, Oreg.: Northwest Regional Educational Lab., October 1975. 25p. ED 127 342.

The basic thesis of the larger paper from which this condensation is drawn is that the use of evaluation contracts is advisable when evaluations are to be conducted by persons external to the institution responsible for the program to be evaluated (or, in larger institutions such as large universities, by persons external to the unit or department responsible for the program). The rationale presented in the later sections has led the authors to propose a set of criteria to assist administrators and evaluators as they think about whether to set up an evaluation contract and, if so, how to go about it. These criteria are summarized in this paper in the form of a checklist. The checklist contains seven subsections which deal respectively with the following types of criteria: (1) criteria for determining whether to conduct an evaluation; (2) criteria for determining whether to contract with an external contractor; (3) criteria to consider when selecting an evaluator; (4) criteria for selecting among procedural options for letting a contract; (5) criteria to consider when using a request for proposals; (6) criteria for use in neogtiating the contract; and (7) criteria for use in monitoring the contract. Different sections of the checklist will be useful to different individuals and groups for different purposes. A flowchart in which the major points of the checklist are translated into a pictorial sequence of events and decisions is also included.

C. PROPRIETY STANDARDS

2. Conflict of Interest. Conflict of interest, frequently unavoidable, should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

Clark, Woodrow W., Jr.; Beers, C. David. Ethical Considerations in the Anthropological Evaluation of Educational Programs. Paper presented at the annual meeting of the American Educational Research Association, 1976. 34p. ED 129 844.

In placing the issue of the ethics of using anthropological methods for educational evaluation in the context of scientism, anthropology and other social sciences are viewed as being in part either qualitative or quantitative. Furthermore, the difference between research and evaluation places the ethnographer in another position in relationship to those studied. Two basic categories of ethical considerations are discussed: data gathering, including loyalty and employment of the field worker, methodology, and confidentiality; and the results of data collection, including the right of review, dissemination of findings, and impact of the data. In Section B, entitled "The Interaction of Ethics and Method," some of the ethical issues involved in designing a research methodology are dealt with. Comments are based on the experience of studying Project Follow Through using a group interview technique.

Molner, Stanley F. Trapped Bedfellows: A Comment on Windle and Neigher. Evaluation and Program Planning, v1 n2 p109-112, 1978. EJ 191 615.

Molner favorably reviews Windle and Neigher's (1978) paper on ethical problems in program evaluation, but suggests that these problems are more political than ethical, and that ethical choices cannot be compromised whereas political choices can.

Scheirer, Mary Ann. Program Participants' Positive Perceptions: Psychological Conflict of Interest in Social Program Evaluation. Evaluation Quarterly, v2 n1 p53-70, February 1978.

A common dilemma of evaluation researchers, that outcome findings do not confirm program administrators' and recipients' perceptions of benefits occurring, is related to a general proposition that participants will have positive perceptions of program effects, regardless of behavioral changes toward program goals. This phenomenon is shown to occur widely, and to be predictable from both behavioral and cognitive social psychological theory, but has not been previously recognized explicitly. Implications are drawn for the policy planning process and for the methodology of program evaluation.

Scriven, Michael. Evaluation Bias and Its Control. In Glass, Gene V. (Ed.), Evaluation Studies Review Annual. Volume 1. Beverly Hills, Calif.: SAGE Pub., 1976. Chapter 5. (Also available as ED 164 593.)

The problem of obtaining unbiased information about the merits of a program or product is considered. Some typical cases of bias include divided loyalty and the co-option of staff evaluation, and divided loyalty and project monitoring. Two principles are helpful in minimizing bias: (1) no unit should rely entirely on a given subunit for evaluative feedback about that same subunit; and (2) since independence is very unstable in an organizational structure, provision must be made to insure and continually reinsure the independence of the evaluators. Four approaches are capable of upgrading the objectivity of evaluation: (1) standardization or routinization of qualitative aspects of the procedures; (2) upgrading the training procedures for evaluators; (3) using the methodology of goal-free evaluation; or (4) using an advocate team approach.

Sheinfeld, Sherri Nita. The Evaluation Profession in Pursuit of Value. Evaluation and Program Planning, v1 n2 p113-115, 1978. EJ 191 616.

Six values useful for judging the ethical problems in program evaluation are: (1) distributive justice; (2) truth seeking; (3) human dignity; (4) sharing; (5) concern for the quality of life; and (6) client loyalty. This paper is a comment on Windle and Neigher (1978).

Sieber, Joan E.; Sanders, Nancy. Ethical Problems in Program Evaluation: Roles, Not Models. Evaluation and Program Planning, v1 n2 p117-120, 1978. EJ 191 617.

The evaluator must begin with a clear understanding of roles, issues, and risks in order to minimize the pressures and occurrence of ethical conflicts in program evaluation. A list of such issues is included in this review of Windle and Neigher (1978).

Windle, Charles; Neigher, William. Ethical Problems in Program Evaluation: Advice for Trapped Evaluators. Evaluation and Program Planning, v1 n2 p97-107, 1978. EJ 191 614.

Ethical problems in program evaluation are increased when conflicting or incompatible models are applied concurrently. Three models are illustrated: an amelioration model, for a program's own decision-makers; an accountability model, focusing on public data disclosure; and an advocacy model, designed to advance the program's interest. Case examples are presented of each. Evaluators should consider several general activities to prevent or solve ethical problems: clarify roles; build organizational supports; be humble; "no fault" program evaluation; give priority to the amelioration model; develop better understanding of ethical aspects of program evaluation; and design legal supports.

C PROPRIETY STANDARDS

3. Full and Frank Disclosure. Oral and written reports should be open, direct, and honest in their disclosure of pertinent findings, including the limitations of the evaluation.

Stake, Robert E. Evaluating Educational Programmes: The Need and the Response. Paris: Organization for Economic Cooperation and Development; Centre for Educational Research and Innovation, 1976. 94p. ED 142 565. (Paper copy available only from organization for Economic Cooperation and Development Publications Center, Suite 1207, 1750 Pennsylvania Avenue, NW., Washington, D.C. 20006, $4.50.)

This survey of recent developments in educational program evaluation is intended for persons who commission, implement, direct, or carry out evaluation studies. The attitudes of government officials, educators, and researchers toward assessment and their own evaluation needs are discussed. Various approaches to evaluation are briefly described; the author emphasizes informal methods as opposed to standard psychometric measures. Instructions for estimating costs of an evaluation project are not provided, but suggestions for effective use of funds are included. Advice is given for evaluators in planning an evaluation study and negotiating an agreement with the monitoring officials. Hypothetical conversations between an official and a prospective evaluator are included as examples that will help the reader start an evaluation properly.

C. PROPRIETY STANDARDS

4. Public's Right to Know. The formal parties to an evaluation should respect and assure the public's right to know, within the limits of other related principles and statutes, such as those dealing with public safety and the right to privacy.

Gooler, Dennis D. Evaluation and the Public. In House, Ernest R. (Ed.), School Evaluation: The Politics and Process. Berkeley: McCutchan, 1973. Chapter 23.

The major assumptions of this chapter are: (1) it is possible and desirable for the public to understand more thoroughly what goes on in fomal educating institutions; (2) it is possible and desirable for the public to have clear and reasonable access to the policy-making processes; (3) increased involvement of people in the organizing and implementing of educational endeavors is in itself desirable, even if inefficiency should ensue; (4) some people would like to be more involved in their schools, if they knew how; and (5) evaluation practices might be useful in accomplishing these purposes. It is proposed that a public education information agency be established to facilitate two-way information exchange, and to monitor the subsequent implications of this exchange for both schools and multiple publics. The agency's role would be to aid the public in understanding its own needs in relation to the school's program, to aide the schools in dispensing information that is helpful to various public groups, and to aid the schools in interpreting responses from the public, as well as helping them assess its various priorities.

C. PROPRIETY STANDARDS

5. Rights of Human Subjects. Evaluations should be designed and conducted, so that the rights and welfare of the human subjects are respected and protected.

Baumrind, Diana. Some Thoughts on Ethics of Research: After Reading Milgram's "Behavioral Study of Obedience." American Psychologist, v19 n6 p421-423, June 1964.

Certain problems in psychological research require the experimenter to balance career and scientific interests against the interests of prospective subjects. Where experimental conditions expose the subject to loss of dignity, or offer nothing of value to the subject, the experimenter is obliged to consider the reasons why the subject volunteered and to offer appropriate rewards. The experimental objectives of the psychologist are seldom incompatible with the subject's ongoing state of well-being, provided that the experimenter is willing to take the subject's motives and interests into consideration when planning the experiment.

Ethical Standards of Psychologists. Washington, D.C.: American Psychological Association, 1953, 1963, 1965, 1972.

Specific principles are addressed to responsibility, competence, moral and legal standards, misrepresentation, public statements, confidentiality, client welfare, client relationship, impersonal services, announcement of services, interprofessional relations, remuneration, test security, test interpretation, test publications, research precautions, publication credit, responsibility toward organization, and promotional activities.

Final Regulations Amending Basic HHS Policy for the Protection of Human Research Subjects. Federal Register, v46 n16 p8366-8392, January 26, 1981.

The Department of Health and Human Services amended the policy for the protection of human research subjects, substantially reducing the scope of the existing regulatory coverage by exempting broad categories of research which normally present little or no risk of harm to subjects.

Kelman, Herbert C. The Rights of the Subject in Social Research: An Analysis in Terms of Relative Power and Legitmacy. American Psychologist, v27 n11 p989-1016, November 1972.

The increasing use of social research in American society and its increasing relevance to public policy and social decisions have engendered widespread concerns about the ethical implications of such research activities. These concerns are of two kinds: (1) concerns relating to the processes of social research, which are exemplified by the issue of invasion of privacy and its various ramifications; and (2) concerns relating to the products of social research, which focus largely on the fear that social research may provide tools for controlling and manipulating human behaviors. The ethical problems surrounding social research can be conceptualized in terms of the power relationship between the subjects and the scientist or user of the research. These problems should be dealt with by overcoming or counteracting the subject's power deficiency.

Shiffer, Lois J. Legal Issues Regarding Sex Bias in the Selection and Use of Career Interest Inventories. In Tittle, Carol Kehr; Zytowski, D. G. (Eds.), Sex-Fair Interest Measurement: Research and Implications. Washington, D.C.: National Institute of Education, 1978. Pages 135-47.

This paper sets forth the various sources of law which set requirements on guidance test selection and use for school systems and counselors. It focuses on laws relevant to test use for career guidance, and indicates steps which counselors and teachers can take to select among and use currently available tests in a manner which complies with legal requirements. It indicates what interpretative materials are available from test publishers, and what materials should be made available directly to students. Finally, it sets forth suggestions for guarding against bias in the use of career inventories.

Weinberger, JoAnn; Michael, John A. Federal Restrictions on Educational Research: A Status Report on the Privacy Act. Educational Researcher, v6 n2 p5-8, February 1977. EJ 156 146.

In a discussion of the Privacy Act of 1974, this article notes that it establishes minimal standards for the protection of individual privacy. By contrast, educational researchers and the social science community generally have a far more restrictive attitude toward the protection of individually identifiable data pertaining to research subjects.

Weinberger, JoAnn; Michael, John. Federal Restrictions on Educational Research: A Status Report on the Buckley Amendment and Freedom of Information Act. Educational Researcher, v5 n11 p3-8, December 1976.

This article summarizes the major actions taken by the Federal government regarding individual privacy and freedom of information, comments on their nature and impact, and highlights current and pending developments.

C. PROPRIETY STANDARDS

6. Human Interactions. Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation.

Everhart, Robert B. Between Stranger and Friend: Some Consequences of "Long Term" Fieldwork in Schools. American Educational Research Journal, v14 n1 p1-15, Winter 1977. EJ 168 848.

Some major consequences of doing fieldwork in schools over an extended period of time are described. Using Powdermaker's distinction of "stranger and friend," the balance between these two roles and perspectives in terms of role, reciprocity, and receptivity is traced through a description of the author's two year study of student life in a junior high school. The paper first describes the evolution from stranger to friend by examining role relationships between the fieldworker and his or her informants. It then focuses upon the problem of the fieldworker having to take on some of the characteristics of the groups being studied. Finally the paper discusses both the beneficial and counter-productive tendencies of these positions for the fieldworker's receptivity to insights about a "familiar" setting. Conclusions center around the dynamic interaction between stranger and friend in long-term fieldwork in educational settings.

Joyce, John F. Humanistic Education Through an Analysis of Evaluation Practices. Journal of Education, v157 n3 p39-51, August 1975. EJ 125 152.

An analysis of the content, process, and purposes of common evaluation practices has revealed several specific dehumanizing effects on participating students and educators. More humanistic, alternative evaluation practices are suggested for each.

Rodman, Hyman; Kolodny, Ralph L. Organizational Strains in the Researcher-Practitioner Relationship. Human Organization, v23 n2 p171-182, 1964. (Also in Caro, Francis G. (Ed.), Readings in Evaluation Research. Second Edition. New York: Russell Sage Foundation, 1977. Also in Gouldner, Alvin; Miller, S. M. (Ed.), Applied Sociology: Opportunities and Problems. New York: Free Press, 1965.)

Potential conflicts between researchers and practitioners are reviewed. They include the evaluative role of the researcher, the differences in the way they organize this time, credit and anonymity, patterns of communication, and the relationship between the researcher and the administrator. Responses to resultant strains include denial and displacement, one-way humor, and various organizational responses.

Ulschak, Francis L.; Weiss, Roland G. The Interpersonal Aspects of Evaluation: A Transactional Analysis Model for Viewing Evaluator-Client Relationships. Educational Technology, v16 n11 p18-25, November 1976. EJ 148 543.

While it is often recognized that interpersonal problems can be a source of difficulties for the evaluator, there seems to be a lack of tools available to aid the evaluator in understanding and dealing with such problems. The purpose of this article is to introduce Transactional Analysis (TA) and propose it as an explicit and practical model which fills this need.

Weiss, Carol H. The Turbulent Setting of the Action Program. In Weiss, Carol H., Evaluation Research: Methods for Assessing Program Effectiveness. Englewood Cliffs, N.J.: Prentice-Hall, 1972. Chapter 5.

Action programs frequently change and evolve during the period under study. Though the evaluator cannot usually control these changes he or she can document and analyze any significant changes. Relationships with program personnel can also cause friction. Possible sources of friction include personality differences; differences in role; lack of clear role definition; conflicting goals, values, interests, frames of reference, or institutional characteristics. Issues that can lead to friction include data collection; changes in record-keeping procedures; selection of program participants; control groups; feedback of information into the program; or status rivalry. Certain conditions appear to be successful in enabling people to function together comfortably: support from administrators; involvement of practitioners in the evaluation; minimizing disruptions; emphasis on theory; the feedback of useful information; and clear role definitions and authority structure.

C. PROPRIETY STANDARDS

7. Balanced Reporting. The evaluation should be complete and fair in its presentation of strengths and weaknesses of the object under investigation, so that strengths can be built upon and problem areas addressed.

Preparing Evaluation Reports: A Guide for Authors. Washington, D.C.: Office of Education (DHEW), 1970. 74p. ED 047 002.

This guide discusses in detail a variety of issues important to the preparation of a good evaluation report. Main sections are concerned with describing the context of the program (locale, school system, etc.); explaining the program (scope, personnel, procedures, etc.); reporting the evaluation (objectives, sample, measuring and reporting change, analysis and presentation of data, etc.); preparing recommendations; and writing the summary. In each section relevant questions referring to matters which should be considered are asked and answered, accompanied in many instances by short example narratives. Additional aid is provided in the form of reference lists of standard works, ordered by difficulty level, on a variety of topics: research methodology, sampling, test theory and construction, and data analysis. The guide concludes with an example of a complete narrative report.

C. PROPRIETY STANDARDS

8. Fiscal Responsibility. The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible.

Sladek, Frea E.; Stein, Eugene L. Grants Budgeting and Finance: Getting the Most Out of Your Grant Dollar. New York:. Plenum Pub., 1981.

This text on the management of grant money covers the entire grant spending process from the decision to apply for a grant to the auditor's final approval. It addresses such issues as applying for a grant, negotiating the best deal, monitoring the fiscal and technical progress of a project, tips for cost sharing, funding agency contracts, and cash management.

D. ACCURACY STANDARDS

1. Object Identification. The object of the evaluation (program, project, material) should be sufficiently examined, so that the form(s) of the object being considered in the evaluation can be clearly defined.

Fullan, Michael; Pomfret, Alan. Research on Curriculum and Instruction Implementation. Review of Educational Research, v47 n2 p335-97, Spring 1977. EJ 166 914

Implementation is not simply an extension of planning and adoption processes; it is a phenomenon in its own right. The main purpose of this review is to explicate the meaning of implementation and its potential determinants by identifying and critically assessing research evidence on the process of curriculum and organizational implementation in schools.

Leinhardt, Gaea. Modeling and Measuring Educational Treatment in Evaluation. Review of Educational Research, v50 n3 p393-420, Fall 1980. EJ 239 573.

The growth of educational evaluation has brought with it a corresponding increase in the desire and need to include information on the nature of the educational treatment that is to be evaluated. However, to date, there has been no systematic review of how this might be accomplished. This paper explores ways in which treatment can be described: either by means of estimating degree of implementation, or by modeling the instructional domain. The paper also reviews approaches to measuring aspects of the instructional environment that are suggested by the various models and methods for combining and analyzing those measures.

Leithwood, Kenneth A.; Montgomery, Deborah J. Evaluating Program Implementation. Evaluation Review, v4 n2 p193-214, April 1980. EJ 222 671.

A methodology for evaluating program implementation is described. Requirements for such a methodology are derived from an analysis of the functions to be performed by implementation evaluation, the nature of the program being implemented, and characteristics of the implementation process. Central features of the methodology involve procedures for the development of a multidimensional profile of the program as it evolves in practice from non- to full implementation. The profile then serves as the basis for instrument development; data collected through the instruments locate program user behavior in relation to the dimensions and levels of use described by the profile. Uses of resulting data to serve program management goals are outlined.

Patton, Michael Quinn. Focusing the Evaluation Question. In Patton, Michael Quinn, Utilization-Focused Evaluation. Beverly Hills, Calif.: SAGE Pub., 1975. Chapter 5.

Once relevant decision-makers and information-users have been identified and organized, the second step in utilization-focused evaluation is to identify and focus the relevant evaluation question. From a utilization point of view, the right evaluation question has several characteristics: (1) it is possible to bring data to bear on the question; (2) there is more than one possible answer to the question; (3) the identified decision-makers want information to help answer the question; (4) they feel they need information to help them answer the question; (5) they want to answer the question for themselves, not just for someone else; (6) they care about the answer to the question; and (7) they can indicate how they would use the answer to the question.

Sjogren, Douglas D. Measurement Techniques in Evaluation. Review of Educational Research, v40 n2 p301-320, 1970.

The increased comprehensiveness of evaluation efforts and a recognition of what is being evaluated has required an expansion of the number and type of measurements included in the evaluation. Observation systems, interaction analysis, matrix sampling, generalizability theory, computer-controlled testing, and mastery testing all have important potential as techniques to measure inputs, processes, and outcomes.

Steinmetz, Andres. Program Evaluation vs. Program Improvement and Some Implications for Training Evaluators. Paper presented at the annual meeting of the American Educational Research Association, 1976. 33p. ED 128 470.

To evaluate a specific program means to compare it against a standard that specifies what the program should be like at a specific time. Standards may be constructed in three ways: surface standards force the definition into a model shaped by scientific procedure; deep standards include information on the inputs, processes, and outputs for each component and subcomponent; and profound standards reach deeper into the organization than its task structure and exhaustively covers all dimensions of organizational functioning, and are set by the program staff. In this situation then, the role of the evaluator includes expressing and explicating the standards set by the program staff, and confronting management with the decisions they must make. The evaluator must also look at the broader environment of the educational organization, i.e., the sociotechnical systems of which schools are a part. This broader perspective implies that the ability to build models; the ability to be able to gather data relative to a large variety of different phenomena; and to report these data using print, verbal, nonverbal, visual and auditory media be included as part of an evaluator's training.

D. ACCURACY STANDARDS

2. Context Analysis. The context in which the program, project, or material exists should be examined in enough detail, so that its likely influences on the object can be identified.

Denny, Terry. Story Telling and Educational Understanding. Paper #12 in Occasional Paper Series. Kalamazoo: School of Education Western Michigan Univ., November 1978. 29p. ED 170 314. (Paper copy available only from the Evaluation Center, Western Michigan University, Kalamazoo, MI 49008.)

Story telling is defined as a kind of journalistic documentation, based on directly observable referents, and used to contribute to an understanding of educational problems. In the area of educational research, story telling is part of the genre which includes case studies, ethnography, and ethnology. Story telling describes an environment in order to communicate a general understanding of a situation. Fieldwork is the essential ingredient in story telling. In educational research, fieldwork consists of familiarizing oneself with local institutions and organizations; talking to local officials and citizens; and particularly, communicating with school personnel and working in the school. The keys to successful fieldwork are the ability to listen; the mastery of interviewing techniques; acute observation skills; and facility at synthesizing information. Story telling, as an ethnographic approach, shows what is happening but does not necessarily reveal causes, and may not be the proper evaluation method if the purpose of the study is to prescribe change or to determine policy decisions.

Scheyer, Patricia T.; Stake, Robert F. A Program's Self-Evaluation Portfolio. Studies in Educational Evaluation, v2 n1 p37-40, Spring 1976. EJ 168 884.

Though responsive evaluation procedures may appear formidable, this paper suggests a way of organizing self-evaluation for projects with only a small budget of money and time. The idea is to establish a file or collection of records or materials which broadly represent the program. This portfolio should be a loose collection so that parts of it can be differently displayed from time to time. The entries should reflect the program activities, its issues, its valuings, and it compromises. The purpose of the portfolio is to aid and broaden out the ordinary evaluation efforts of the program staff.

D. ACCURACY STANDARDS

3. Described Purposes and Procedures. The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.

-

Mager, R. F. Goal Analysis. Belmont, Calif: Fearon Pub., 1972.

This book is designed to teach the reader to identify statements that describe abstractions and those that describe performances; and, after identifying all important goals, to be able to describe the performances that represent achievement of the goals.

Sanders, James R.; Nafziger, Dean H. A Basis for Determining the Adequacy of Evaluation Designs. Portland, Oreg.: Northwest Regional Educational Lab., October 1975. 57p. ED 127 345.

A basis is provided for judging the adequacy of evaluation plans or evaluation designs in this document. It is assumed that using the procedures suggested to determine the adequacy of evaluation designs in advance of actually conducting evaluations will lead to better evaluation designs, better evaluations, and more useful evaluative information. The paper is divided into four general sections. First, some basic questions are considered--Why evaluate? Why do we need evaluation designs? Why do we need a basis for judging the adequacy of an evaluation design? Answers to these questions serve to underscore the importance of providing a consistent basis for judging evaluation designs. Second, a checklist of basic considerations important in judging evaluation designs is presented. Third, a sample design is presented, together with an example of how the checklist can be used in judging a design. Fourth, professional educators' thoughts about judging the adequacy of evaluation designs are presented.

Stufflebeam, Daniel L. Meta-Evaluation. Paper #3 in Occasional Paper Series. Kalamazoo: School of Education, Western Michigan University, December 1974.

Good evaluation requires that evaluation efforts themselves be evaluated. Many things can and often do go wrong in evaluation work. Accordingly, it is necessary to check evaluations for problems such as bias, technical error, administrative difficulties and misuse. Such checks are needed both to improve ongoing evaluation activities and to assess the merits of completed evaluation efforts. This paper presents both a logical structure and methodological suggestions for evaluating evaluations. Part I analyzes background factors and problems associated with meta-evaluation, the need for metaevaluation, and summarizes pertinent literature. Suggestions are made concerning what criteria should guide the development of meta-evaluation methodology. Finally, six classes of problems that jeopardize evaluation and need to be addressed by meta-evaluation methodology are enumerated. Part II is a conceptual response to the first part. It defines and sets forth premises for meta-evaluation and presents a logical structure for designing meta-evaluation studies. Part III applies the logical structure presented in the previous section. It contains five meta-evaluation designs, four for use in guiding evaluation work, and the fifth for judging completed evaluation work.

D. ACCURACY STANDARDS

4. Defensible Information Sources. The sources of information should be described in enough detail, so that the adequacy of the information can be assessed.

Note: See also the references for Standards D5 and D6.

Campbell, Donald T. Keeping the Data Honest in the Experimenting Society. In Moulton, H. W.; Watson, D. J. H., Interdisciplinary Dimensions of Accounting for Social Goals and Social Organizations. Columbus, Ohio: Grid, 1977.

The "experimenting society" is proposed as an alternative future. It would be scientific, nondogmatic, honest, accountable, and challengeable. There are, however, many methodological problems to be solved before this society can be implemented, e.g., the issues of randomized experiments, opinion surveys, social indicators, and use of multiple indicators. The ensuing discussion of related issues groups them as metascientific issues, statistical issues, and political system problems.

Cochran, Nancy. Grandma Moses and the "Corruption" of Data. Evaluation Quarterly, v2 n3 p363-73, August 1978. EJ 186 186.

Distortion of data is caused by purposeful, goal-oriented activity of people who produce data, as well as by attempts to cheat or manipulate social service delivery systems. Failure to recognize a constructive motivational component is attributed to an over-reliance on positivism in the social sciences. It is argued that increased regulation may actually increase distortion and decrease the availability of valid information about social services. Legitimization, self-knowledge, and understanding dynamic processes are suggested as alternatives to using program evaluation for measuring effects of social intervention.

David, Jane L.; Relavin, Sol H. Evaluating Compensatory Education: Over What Period of Time Should Achievement Be Measured? Journal of Educational Measurement, v15 n2 p91-99, Summer 1978. EJ 189 635.

The goal of compensatory education to increase achievement implies that some of this increase should be sustained beyond the end of the program. This paper presents data that allow comparisons between the traditional fall-to-spring evaluation period and a fall-to-fall time period. Analysis show that students in compensatory programs often suffer substantial losses in achievement over the summer. Therefore, fall-to-fall achievement gains are smaller than the traditional fall-to-spring gains. This difference in gains can lead to very different conclusions about a program's success. If the goal is sustained achievement, evaluations should be based, at a minimum, on a fall-to-fall time period.

Poynor, Hugh. Selecting Units of Analysis. In Borich, Gary D. (Ed.), Evaluating Educational Programs and Products. Englewood Cliffs, N.J.: Educational Technology Pub., 1974. Chapter 15.

A proper unit of analysis is the smallest source of data that is both logically and statistically defensible. Both approaches separate pupil units from classroom averages, although to different degrees. Simulated empirical demonstrations are used to reveal the importance of choosing the proper unit of analysis.

Sawin, Enoch I. Curriculum Evaluation or Descriptive Inquiry. Studies in Educational Evaluation, v2 n1 p41-51, Spring 1976. EJ 168 885.

Problems associated with current expertise in evaluation are discussed. Since evaluators are not always able to reliably achieve all levels of an evaluation project, these tasks are categorized into five levels of complexity. The author suggests that evaluators should retrench down the scale of complexity of an evaluation until a level is reached at which (1) conclusions are reliable across investigators and are scientifically defensible, (2) results obtained pose minimum threats to personnel, and (3) training requirements for evaluators are within reason. A more accurate label for such evaluators would be "descriptive inquiry specialists."

Stake, Robert E. Objectives, Priorities, and Other Judgment Data. Review of Educational Research, v40 n2 p181-212, April 1970.

Four kinds of data are considered judgment data: personal value-commitments, objectives, the priorities given to certain objectives, and standards. In evaluation studies, judgment data should be gathered and analyzed. Surveys, scaling, the Q-technique, the semantic differential, observation, and expert review are all methods for gathering judgment data. Though difficult to summarize, judgment data should be reported, possibly in narrative form or using a profile or matrices. Finally, the jugment data should enter into decision processes as inputs, not as outputs.

Tittle, Carol Kehr. Test Bias: Current Methodology and Implications for Evaluators. In Abramson, Theodore; And Others (Eds.), Handbook of Vocational Education Evaluation. Beverly Hills, Calif.: SAGE Pub., 1979. Chapter 20.

After describing key definitions and requirements in the Uniform Guidelines on Employee Selection Procedures, this chapter reviews the procedures and methods that have been used for examining test and item bias in the educational assessment setting, in the absence of an external criterion. Finally it presents a series of recommendations to evaluators, listing the data that evaluators should find in test manuals, the data that are needed to make the determination that a test is fair for use with particular groups, and the procedures evaluators will find useful in minimizing test bias in local test development.

Tittle, Carol Kehr. Use of Judgmental Methods in Item Bias Studies. In Berk, R. A. (Ed.), Handbook of Methods for Detecting Test Bias. Baltimore: Johns Hopkins Univ. Press, in press.

The renewal of interest in item bias and the fairness of tests used in evaluation has focused attention on the test development process and construct validity. Judgmental methods used throughout the test development process include procedures to examine stereotyping and fair representation of groups. Judgments also provide validity-related evidence: familiarity of groups with the nominal content of items and the opportunity to learn item content and process (the match or overlap of items with the curriculum and the instructional process). Research and procedures in these areas are described.

D. ACCURACY STANDARDS

5. Valid Measurement. The information gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the interpretation arrived at is valid for the given use.

Note: See also the reference for Standards D4 and D6.

Berk, Ronald A. (Ed.) Criterion-Referenced Measurement: The State of Art. Baltimore, Md.: Johns Hopkins Univ. Press, 1980.

This book is a product of the first annual Johns Hopkins University National Symposium on Educational Research, held in Washington, D.C., in October 1978. It attempts to determine the state of the art of criterion-referenced measurement. It includes discussions of content domain specification and item generation, item and test validity, and reliability.

Cook, Thomas. D.; Campbell, Donald T. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand-McNally, 1979.

This book presents some quasi-experimental designs and design features that can be used in many social research settings. Each design is assessed in terms of four types of validity, with special stress on internal validity. Although general conclusions are drawn about the strengths and limitations of each design, emphasis is also placed on the fact that the relevant threats to valid inference are specific to each research setting.

Cronbach, Lee J.; And Others. The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles. New York: John Wiley & Sons, 1972.

This monograph presents a theory for evaluating the generalizability of test scores and profiles, and scores derived from field observations. It contains concrete examples and problems for advanced students in measurement theory and research methodology.

Cronbach, Lee J. Validity on Parole: How Can We Go Straight? New Directions for Testing and Measurement, n5 p99-108, 1980.

Presented at the 1979 Educational Testing Service Invitational Conference, this article reviews developments in test validation in the past decade. As with a scientific theory, interpretation of a test is going to remain open and unsettled, the more so because of the role values play in legal and policy actions based on tests.

Hambleton, Ronald K.; Eignor, Daniel R. Guidelines for Evaluating Criterion-Referenced Tests and Test Manuals. Journal of Educational Measurement, v15 n4 p321-327, Winter 1978. EJ 198 850

A set of guidelines for evaluating criterion-referenced tests is presented. The guidelines address objectives, test items, administration, test layout, reliability, cut-off scores, validity, norms, reporting of test score information, and test score interpretations. Additionally, 11 sets of extant criterion-referenced tests are evaluated using these guidelines.

Messick, Samuel. Test Validity and the Ethics of Assessment. American Psychologist, v35 n11 p1012-27, November 1980. EJ 235 612.

Questions of the adequacy of a test as a measure of the characteristic it is interpreted to assess are answerable on scientific grounds by appraising psychometric evidence, especially construct validity. Questions of the appropriateness of test use in proposed applications are answerable on ethical grounds by appraising potential social consequences of the testing. The first set of answers provides an evidential basis for test interpretation, and the second set provides a consequential basis for test use. By then considering both the evidential and consequential bases of both test interpretation and test use, the roles of evidence and social values in the overall validation process are illuminated, and test validity comes to be based on ethical as well as evidential grounds.

Nunnally, Jum C.; Durham, Robert L. Validity, Reliability and Special Problems of Measurement in Evaluation Research. In Struening, Elmer L.; Guttentag, Marcia (Eds.), Handbook of Evaluation Research. Volume 1. Beverly Hills, Calif.: SAGE Pub., 1975. Chapter 10.

This chapter discusses methods for determining the validity of measures, and principles concerning reliability. In one way or another, all the issues discussed in this chapter concern generalizability. Thus, the validity of a predictor test concerns the extent to which one can generalize from scores on the test to scores on a criterion variable. Reliability concerns the extent to which one can generalize from scores on a test to scores on alternative forms of the test.

Porter, Andrew C.; And Others. Impact on What?: The Importance of the Content Covered. Research Series No. 2. East Lansing: Michigan State Univ., Inst. for Research on Teaching, February 1978. 37p. ED 155 215.

Defining practical significance in program evaluations is a difficult measurement problem which can only be solved by an intimate familiarity with the measures on which effects are estimated and their content relationship to the program goals. Past attempts to provide general solutions to the size of effect problems have relied on standardized indices which can be estimated and reported without any knowledge of what was measured. Such efforts are viewed here as steps in the wrong direction. Instead, what is called for is a procedure whereby the content goals of the program, the content implied by a test, and the interrelationship between the two are made explicit. The procedure should investigate treatment-by-item interactions and at the same time, describe the measures used so that persons other than the evaluator can reach their own decisions about practical significance. Analysis of the mathematics sections of four major intermediate level standardized tests with their taxonomies indicated rather substantial differences in content tested. It was clear that standardized tests are not well suited to the task of estimating item domain by treatment interactions.

Shepard, Lorrie. Purposes of Assessment. Studies in Educational Evaluation, v5 n1 p13-26, 1979. EJ 210 291.

Assessment generally refers to large-scale, system-wide measurement programs for pupil diagnosis; pupil certification; program evaluation; research; accountability; resource allocations; or teacher evaluation. The purpose of assessment should determine the test content, construction, administration, and examinees sampled. Assessment methods for one purpose may be inappropriate for other applications.

Standards for Educational and Psychological Tests. Revised Edition. Washington, D.C.: American Psychological Association, 1974. Also relevant for Standard D6, Reliable Measurement.)

This document presents standards for test use as well as for test manuals; it is intended to guide both test developers and test users. These standards apply to any assessment procedure, assessment device, or assessment aid. They are grouped in three levels: Essential, Very Desirable, and Desirable. The standards cover tests, manuals, and reports; reliability and validity; and the use of tests. They were prepared by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education.

Walker, Clinton B. Standards for Evaluating Criterion-Referenced Tests. Los Angeles: Center for the Study of Evaluation, Univ. of California - Los Angeles, January 1978. 33p. ED 179 595.

Standards for evaluating criterion-referenced tests are presented. Twenty-one standards, grouped in three categories, are discussed. Category One is defined as Measurement Properties and is comprised of conceptual validity, including description of the domain, test item agreement with objectives, and item representativeness of the objectives; and field test validity, including sensitivity, item uniformity, divergent validity, lack of bias, and consistency of scores. Category Two is labelled Appropriateness for Examinees, and is comprised of clarity of instruction; item review; physical format, including layout and legibility; and ease in recording answers. Category Three is called Practicality, and is composed of adequacy of information about the test; relevance of items of at least two series of teaching materials; flexibility, including multilevel testing of objectives; alternative test forms; clarity of test admistration directions; scoring; record keeping; availability of rules to make instructional decisions based on test results; and comparative data on test scores. It is also stated that the test buyer must determine the degree of correspondence between the objectives of a test package and the objectives of the curriculum to be tested.

Wargo, Michael J.; Green, Donald Ross (Eds.) Achievement Testing of Disadvantaged and Minority Students for Educational Program Evaluation. New York: CTB/McGraw-Hill, 1977.

This book represents the proceedings of a conference of the same title held in Reston, Virginia, in May 1976. The purpose of the conference was to identify, define, and analyze problems associated with the use of standardized achievement tests on populations of disadvantaged and minority students for educational program evaluation.

D. ACCURACY STANDARDS

6. Reliable Measurement. The information-gathering instruments and procedures should be chosen or developed and then implemented in ways that will assure that the information obtained is sufficiently reliable for the intended use.

Note: See also the references for Standards D4 and D5.

Cook, Thomas D.; Campbell, Donald T. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally, 1979.

This book presents some quasi-experimental designs and design features that can be used in many social research settings. The designs serve to probe causal hypotheses about a wide variety of substantive issues in both basic and applied research. Each design is assessed in terms of four types of validity--statistical conclusion validity, internal validity, construct validity and external validity--with special stress on internal validity. General conclusions are drawn about strengths and limitations of each design; however, emphasis is also placed on the fact that the relevant threats to valid inference are specific to each research setting. Several chapters deal with quasi-experimental designs and modes of analyzing data that result from them. Another chapter deals with causal inference from designs that lack most of the characteristic features of experimental research. The final chapter states that randomized experiments are sometimes possible in field research, and outlines obstacles to their implementation, some ways of overcoming these obstacles, and some ways of recognizing the situations when random assignment is most feasible.

Cronbach, Lee. J.; Gleser, Goldine C. Interpretation of Reliability and Validity Coefficients: Remarks on a Paper by Lord. Journal of Educational Psychology, v50 n5 p230-237, October 1959.

Most statements describing the usefulness of tests as judged from their reliability or validity coefficients assume that a decision is made about every person tested, i.e., the persons are divided into three classes: those whose true scores are greater than a specified criterion score, those whose true scores are less than the criterion, and those for whom neither interpretation may safely be made. This paper differs from Lord's in placing emphasis upon the maximum risk of erroneous interpretation rather than upon the average risk. The suitability of a test depends not only on the reported reliability and validity coefficients, but also on the importance of the decisions to be made and on the rules by which the scores are to be converted into interpretations.

Linn, Robert L.; Slinde, Jeffrey A. The Determination of the Significance of Change Between Pre- and Posttesting Periods. Review of Educational Research, v47 p121-150, Winter 1977. EJ 161 389.

The major issues that arise in the measurement of change are reviewed and, where possible, alternative approaches are discussed. The measurement of individual differences is considered first. This is followed by a discussion of some of the concerns involved in inferring treatment effects from group differences. The concluding section discusses accountability systems based on student achievement.

Lord, Frederic M. The Utilization of Unreliable Difference Scores. Journal of Educational Psychology, v49 p150-152, June 1958.

The purpose of this paper is to call attention to a natural way in which difference scores having relatively low reliability may be (and currently are) used effectively; and to suggest a method for inferring from the reliability coefficient of difference scores their effectiveness when used as outlined.

Nunnally, Jum C. The Study of Change in Evaluation Research: Principles Concerning Measurement, Experimental Design and Analysis. In Struening, Elmer L.; Guttentag, Marcia (Eds.), Handbook of Evaluation Research. Volume 1. Beverly Hills, Calif.: SAGE Pub., 1975. Chapter 6.

This chapter discusses problems which frequently arise in the study of change in education research. The issues are discussed in the order that they would occur to the scientist: measurement (constructing measures expost facto, reactivity of measurement, faking of responses, extent of measurement problem, subjective assessments), then research design (experimental and quasi-experimental), and finally statistical analysis.

Stanley, Julian C. Reliability. In Thorndike, Robert L. (Ed.), Educational Measurement. Second Edition. Washington, D.C.: American Council on Education, 1971. Chapter 13.

The fact that repeated sets of measurements never exactly duplicate one another is what is meant by unreliability; a tendency toward consistency from one set of measurements to another is called reliability. Methods for estimating reliability are discussed in the context of classical test theory and in light of some more recent approaches to test-score theory.

D. ACCURACY STANDARDS

7. Systematic Data Control. The data collected, processed, and reported in an evaluation should be reviewed and corrected, so that the results of the evaluation will not be flawed.

Ball, Samuel. Audit of Evaluation. In Anderson, Scarvia B.; And Others, Encyclopedia of Educational Evaluation: Concepts and Techniques for Evaluating Education and Training Programs. San Francisco: Jossey-Bass, 1975. Pages 40-42.

The idea of auditing evaluations came from the educational accountability movement that developed in the 1960's. The actual work of the auditor varies somewhat, but can include looking over the evaluation plans, monitoring data collection, checking the analyses, reading an early draft of the evaluation report, suggesting changes in the report, and approving the final version.

Demaline, Randy E.; Quinn, D. William. Hints for Planning and Conducting a Survey and a Bibliography of Survey Methods. Aid #2 in Instructional Aids Series. Kalamazoo: School of Education, Western Michigan Univ., April 1979. 107p. ED 173 417.

Methods of planning and administering mail surveys, developing questionnaires, and analyzing data are reviewed. Each review section is followed by an annotated list of selected readings. Topics discussed in planning a survey include decision-making; survey designs; sampling plans; and ethical considerations. Development of instruments and types of attitude measures are discussed in the section on survey instruments. The section on survey management is concerned with the mechanics of distributing and collecting the questionnaires, coding, and checking for errors. Data analysis focuses on nonresponse analysis, computer usage, and the choice of statistical methods. A 365-item bibliography and a subject index geared to the bibliography are appended.

Murphy, Richard T. Quality Control. In Anderson, Scarvia B.; And Others, Encylopedia of Educational Evaluation: Concepts and Techniques for Evaluating Education and Training Programs. San Francisco: Jossey-Bass, 1975. Pages 299-301.

The methods by which the sample is selected and its quality tested are statistical quality-control methods. For quality control in evaluation, data collected with tests, interviews, observation techniques, ratings, and other methods, must be examined to see whether they are judged sufficiently free of error to be worthy of further analysis. Quality control procedures should be used routinely from the very first phase of data collected through to the last stages of analysis.

Stufflebeam, Daniel L.; And Others. Educational Evaluation and Decision Making. Itasca, Ill.: F. E. Peacock, 1971. Pages 176-197.

When the delineation of information needs is completed, the evaluator must establish a plan to obtain the information. This plan should consist of the following areas: collection of data, organization of data, and analysis of data. These tasks can be further broken down into work units.

D. ACCURACY STANDARDS

8. Analysis of Quantitative Information. Quantitative information in an evaluation should be appropriately and systematically analyzed to ensure supportable interpretations.

Bentler, Peter M.; Woodward, J. Arthur. Nonexperimental Evaluation Research: Contributions of Causal Modeling. In Datta, Lois-ellen; Perloff, Robert (Eds.), Improving Evaluations. Beverly Hills: SAGE Pub., 1979. Chapter 6.

This chapter discusses the relevance of causal modeling research methodologies to evaluation research, reviews in a nontechnical manner a series of causal modeling techniques for both quantitative and qualitative measures, and concludes with an example applying structural equation models to data from a summer Head Start program. This nontechnical introduction is intended as a first step toward assessing causal modeling in evaluation research.

Bryk, Anthony S.; Weisberg, Herbert I. Use of the Nonequivalent Control Group Design When Subjects Are Growing. Psychological Bulletin, v84 n5 p950-962, September 1977.

In the nonequivalent control group design, pretest and posttest data on both groups are obtained. Statistical methods are used to adjust posttest comparisons, based mainly on pretest information. The purpose of this article is to consider the adequacy of these methods, from an individual growth perspective. It is concluded that statistical adjustments are generally inadequate in the face of nonequivalent growth systems across treatment groups.

Cooley, William W.; And Others. Analyzing Multilevel Data. In Berk, Ronald A. (Ed.), Educational Evaluation Methology: The State of the Art. Baltimore: Johns Hopkins Univ. Press, 1981. Chapter 3.

The critical issues in the analysis of multilevel data from evaluation studies are identified and discussed. Their presentation is restricted to explanatory observational studies where the statistical methods usually involve the analysis of relationships among variables, e.g., multiple regression. The review of the issues is organized in four sections: the importance of choosing a causal model prior to choosing a method of analysis; aggregation bias; the implications of the variation that might occur in within-group coefficients; and general strategies for analyzing multilevel data.

Cronbach, Lee J. Analysis of Covariance in Nonrandomized Experiments: Parameters Affecting Bias. Stanford, Calif.: Evaluation Consortium, Stanford University, August 1977.

A model for nonrandom experiments is developed to evaluate the bias in the adjustments made to compare outcomes in nonequivalent groups. The adjustment made in analysis of covariance depends on the covariate employed. The covariate can be expressed as a weighted combination of an ideal covariate, which determines outcome scores within a treatment group; a discriminant, which determines assignment to treatment group; and irrelevant information. The presence of irrelevant information reduces the absolute value of the adjustment. When the covariate contains little or no irrelevant information, the adjustment may be too large or too small, depending on the correlations of the covariate with the discriminant and the ideal covariate. Correction procedures now present in the literature cannot be counted on to provide an unbiased estimate of the treatment effect.

Cronbach, Lee J.; Furby, Lita. How Should We Measure "Change" - Or Should We? Psychological Bulletin, v74 n1 p68-80, July 1970. Errata, Psychological Bulletin, v74 n3 p28, September 1970.

Procedures previously recommended by various authors for the estimation of "change" scores, "residual" or "basefree" measures of change, and other kinds of difference scores are examined. A procedure proposed by Lord is extended to obtain more precise estimates, and an alternative to the Tucker-Damarin-Messick procedure is offered. A consideration of the purposes for which change measures have been sought in the past leads to a series of recommended procedures which solve research and personnel decision problems without estimation of change scores for individuals.

Kenny, David A. A Quasi-Experimental Approach to Assessing Treatment Effects in the Nonequivalent Control Design. Psychological Bulletin, v82 n3 p345-362, May 1975.

Four statistical tests of treatment effect are evaluated for the nonequivalent control group design: analysis of covariance, analysis of covariance with reliability correction, raw change score analysis, and standardized change score analysis. Given a model of the process of selection into treatment groups, the nonequivalent control group design can yield interpretable results.

Marascuilo, Leonard A. Measuring Differences among Non-Randomized Groups: An Epidemiological Model for Identifying Successful School Program. Journal of Experimental Education, v48 n1 p50-59, Fall 1979. EJ 220 353.

It is recommended that the biomedical model of adjusted statistics designed to overcome the difficulty investigators face when attempting to randomize subjects be adopted. The adjusted discrepancies between group statistics are considerably smaller than is indicated by inspection of raw, or unadjusted, sample values. This model provides a way to obtain a more accurate estimate of program success or failure when comparisons across classrooms or other units are desirable.

Porter, Andrew C.; Chibucos, Thomas R. Selecting Analysis Strategies. In Borich, Gary P. (Ed.), Evaluating Educational Programs and Products. Englewood Cliffs, N.J.: Educational Technology Pub., 1974. Chapter 16.

Evaluation paradigms are divided into four categories, determined by the presence or absence of random assignment, and the use of a pretest or the use of some other variable observed antecedent to treatment. For each category of design, the following analysis strategies are considered: analysis of covariance using a random covariate, analysis of variance of an index of response including gain scores as a special case, repeated measures analysis of variance, and analysis of covariance using estimated true scores as the covariate.

Weisberg, Herbert I. Statistical Adjustments and Uncontrolled Studies. Psychological Bulletin, v86 n5 p1149-1164, September 1979.

A variety of problems are related to a lack of experimental control: measurement error, unequal growth rates across groups, and regression artifacts. In this article it is shown that these problems can all be subsumed under a general conceptual framework, as particular examples of model misspecification. The case of linear adjustment (analysis of covariance) is given special attention.

Wolf, Richard M. Selecting Appropriate Statistical Methods. In Berk, Ronald A. (Ed.), Educational Evaluation Methodology: The State of the Art. Baltimore: Johns Hopkins Univ. Press, 1981. Chapter 5.

Several statistical methods for analyzing the results of an evaluation study are compared. Factors that guide the selection of analysis of variance and analysis of covariance are delineated in the first section. The major focus in subsequent sections is the analysis of nonrandomized designs. Special attention is given to designs based on comparable groups and to those based on noncomparable groups.

D. ACCURACY STANDARDS

9. Analysis of Qualitative Information. Qualitative information in an evaluation should be appropriately and systematically analyzed to ensure supportable interpretations.

Alternative Methodology. In Hamilton, David; And Others (Eds.), Beyond the Numbers Game: A Reader in Educational Evaluation. London: Macmillan Education, 1977. Section 4.

Alternative evaluation is an eclectic approach, adaptive and responsive to the particular learning milieu in which the evaluator is working. These are intellectual traditions outside education that illuminative evaluators draw upon: participant observation in sociology, ethnographic field work in social anthropology, literacy criticism, film documentary, historical research, law and clinical psychiatry. A variety of papers are presented in this section, covering case studies, field work and the generation of theory, and the community context of evaluation.

Becker, Howard S. Problems of Inference and Proof in Participant Observation. American Sociological Review, v23 n6 p652-660, December 1958.

The basic analytic operations carried on in participant observation are described. These stages of analysis are conducted in the field: the selection and definition of problems, concepts, and indices; the check on the frequency and distribution of phenomena; and the incorporation of individual findings into a model of the organization under study. A fourth stage of final analysis involves problems of presentation of evidence and proof.

Bogdan, Robert; Taylor, Steven J. Introduction to Qualitative Research Methods: A Phenomenological Approach to the Social Sciences. New York: John Wiley, 1975.

Over the past decade, there has been growing interest in the subjective, in meaning, and in common-sense understandings. This book discusses qualitative methods as they relate to the phenomenological perspective. Part one contains a "how to do it" approach to participant observation, personal documents, open-ended interviews, and examples of qualitative studies. Part two is a discussion of how to present findings, with several example reports.

Everhart, Robert B. Problems of Doing Fieldwork in Educational Evaluation. Human Organization, v34 n2 p205-215, Summer 1975. EJ 119 820.

The use of fieldwork in evaluation presents the fieldworker and agencies connected with educational evaluation with a number of critical problems, four of which are identified and discussed in this paper. The first, the identification of the evaluation problem, notes the distinctions between the traditional evaluator who defines evaluation problems in an a priori manner and the fieldworker who defines the problem holistically and as a result of preliminary fieldwork. The second problem focuses upon the unclear and changing signals which the fieldworker receives from the agency sponsoring the evaluation. A related problem is the multiplicity of expectations the fieldworker receives from various members in his or her role set. The dissemination of data and consequences of dissemination is the third problem area discussed. The paper concludes with an examination of some compromises necessitated by the fieldworker doing evaluation work.

Ferreira, Joseph; Burges, Bill. Collecting Evidence: A Layman's Guide to Participant Observation. Boston: Institute for Responsive Education, 1976. 28p. ED 132 715. (Paper copy available only from the Institute for Responsive Education, 704 Commonwealth Avenue, Boston, MA 02215.)

Participant observation is useful as a tool for gathering evidence about processes, circumstances, or other observable conditions. A participant/observer is an investigator gathering evidence. Observations are carefully recorded, prejudgment is scored, and judgments flow from the evidence. In approaching a situation to be investigated, the participant/observer should get a flavor of the system of which the situation is a part, identify those problems or parts of the system that influence the situation under investigation, and select the problems or parts that seem most important and might provide vital evidence. Once the situation is chosen, five types of data are often important: descriptive data about settings, accurate descriptions of actions and behaviors, word-for-word statements, traces and wear spots, and documents. The participant/observer is also interested in reliable witnesses and informants. Self-training exercises and sample observations are included.

Fienberg, Stephen E. The Collection and Analysis of Ethnographic Data in Educational Research. Anthropology and Education Quarterly, v8 n2 p50-57, May 1977.

The major theme of this paper is that, from a scientific viewpoint, there is no fundamental difference between the two sides of the qualitative/quantitative controversy. The process of statistical inference is basically the same for both types of research. Ethnographic researchers have pinpointed a major flaw in much educational research: the unit of analysis need not be the same as the apparent unit of sampling. Finally, in addition to using multivariate methods to analyze their data, investigators need to begin thinking in terms of large-scale randomized controlled field trials.

Filstead, William J. Using Qualitative Methods in Evaluation Research: An Illustrative Bibliography. Evaluation Review, v5 n2 p259-268, April 1981.

This article briefly describes the ways in which qualitative methods have been viewed relative to evaluation research. The topics included in the bibliography include: the changing climate in evaluation research, the philosophical and conceptual background behind this approach to research, actual evaluation efforts which employed qualitative methods, the use of various data gathering techniques, and how one "makes sense" of these data.

Guba, Egon G. Toward a Methodology of Naturalistic Inquiry in Educational Evaluation. CSE Monograph Series in Evaluation, 8. Los Angeles: Center for the Study of Evaluation, Univ. of California-Los Angeles, 1978. 97p. ED 164 599.

Evaluation is viewed as essential to decision making and social policy development. Since conventional methods have been disappointing or inadequate, naturalistic inquiry (N/I) differs from conventional science in minimizing constraints on antecedent conditions (controls) and on output (dependent variables). N/I is phenomenological rather than positivist. It offers alternative strategies for problems when the experimental approach is implausible. A number of new evaluation models (such as the responsive model, the judicial model, and the connoisseurship model) are compatible with the approach. Since there is no compelling way to truth, N/I must be credible and deal convincingly with standard methodological problems such as boundary problems (setting the scope of inquiry), focussing problems (establishing and defining categories), and problems of authenticity (reliability, validity, and objectivity). Techniques for establishing validity include: triangulation, cross examination, persistent observation, and peer or participant corroboration. In evaluation, validity may be ecological, contextual, or phenomenological. Impartiality is imperilled by conscious or unconscious bias, incompetence, gullibility, or corruptibility. It is promoted by openness and fairness.

Krippendorff, Klaus. Content Analysis: An Introduction to Its Methodology. Beverly Hills, Calif.: SAGE Pub., 1980.

Content analysis is an important research technique in the social sciences. The methodology seeks to understand data not as a collection of physical events, but as symbolic phenomena and to approach their analysis unobtrusively. This book presents three aspects of content analysis: its theory, methods and procedures, and qualitative criteria. In discussing theory, a brief history of content analysis is presented as well as a definition that distinguishes content analysis from other methods and exemplifies its domain of practical applications. Within methods and procedures, the following is presented: the logic of designs, units of analysis, sampling, recording, construction of data languages, analytical constructs, computational techniques, and the use of computers. The quality criteria of content analysis are reliability and validity. Suggestions are made as to how to meet both criteria. Finally, the book concludes with a practical guide for doing content analysis.

Lofland, John. Analyzing Social Settings: A Guide to Qualitative Observation and Analysis. Belmont, Calif.: Wadsworth Pub. Company, 1971.

A positive and detailed set of instructions are presented indicating exactly how qualitative observation and analysis are performed. In the first three chapters there is an attempt to specify what qualitative analysis is and how it differs from quantitative analysis. There follow concrete descriptions of the two basic techniques used by qualitative observers in collecting their materials: interviewing and participant-observation. Finally, recommendations are made as to how one can store and organize materials to facilitate more acute observation, analysis, and writing.

Smith, Louis M. An Evolving Logic of Participant Observation, Educational Ethnography, and Other Case Studies. In Schulman, L. C. (Ed.), Review of Research in Education, v6, 1978.

This chapter provides a context and logic for the discussion of educational ethnography by making three major points. First, a large body of both substantive and methodological literature within this field study tradition already exists. Second, a reflexive overview of the cognitive processes in field work suggests a perspective on methodology. Third, the essay presents a patterned analysis of this genre of research, considering four major domains: data, descriptive narrative, theoretical, and metatheoretical.

Smith, Louis. Integrating Participant Observation into Broader Evaluation Strategies. In Hamilton, David; And Others (Eds.), Beyond the Numbers Game: A Reader in Educational Evaluation. London: Macmillan Education, 1977. Section 4. Chapter 6.

Participant observation can be integrated with other evaluation techniques in three different general evaluation strategies: a general structural model, a sequential model, and a case-study accumulation model. The general structural model brings together three research strategies: an experimental design, a social survey, and participant observation. The sequential model attempts to cumulate efforts over time rather than concurrently in time. The third strategy is the cummulation of participant observer case studies.

Trend, M. G. On the Reconciliation of Qualitative and Quantitative Analyses: A Case Study. In Cook, T.; Reichardt, R. (Eds.), Qualitative and Quantitative Methods in Evaluation Research. Beverly Hills, Calif.: SAGE Pub., 1979. Chapter 4.

This paper examines an instance where the analysis of qualitative data from a participant observer produced an explanation that could not be reconciled immediately with one based upon quantitative data from the same social experiment. The presentation is designed to: (1) give the reader insight into the social psychology which operates in large-scale research efforts; (2) dispel the notion that using multiple methods will lead to sounder explanations in an easy, additive fashion; and (3) suggest a way of proceeding in resolving a conflict between two different explanations of the same events.

Wolcott, Harry. Criteria for an Ethnographic Approach to Research in Schools. Human Organization, v34 n2 p111-128, Summer 1975.

A number of issues related to conducting ethnographic research in schools are explored under four criterion headings: (1) appropriateness of the problem, (2) appropriateness of the ethnographer, (3) appropriateness of the research "climate," and (4) appropriateness of expectations for the completed study. The purpose of the paper is to air some of the issues that arise between those who do ethnography in educational settings and those who commission it, by examining recent experience and customary expectations. With the express purpose of fostering further dialogue, the author states his own position on many of the issues, but the critical function served by the paper is to work toward explicating problems rather than necessarily resolving them.

D. ACCURACY STANDARDS

10. Justified Conclusions. The conclusions reached in an evaluation should be explicitly justified, so that the audiences can assess them.

Note: See also the references for Standards A4 and D8.

Campbell, Donald T.; Erlebacher, Albert. How Regression Artifacts in Quasi-Experimental Evaluations can Mistakenly Make Compensatory Education Look Harmful. In Struening, Elmer L.; Guttentag, Marcia (Eds.), Handbook of Evaluation Research. Volume 1. Beverly Hills: SAGE Pub., 1975. Chapter 19. (Also in Helmuth, J. (Ed.), Compensatory Education: A National Debate, Volume 3 of The Disadvantaged Child. New York: Brunner/Mazel, 1970.)

Evaluations of compensatory-educational efforts such as Head Start are commonly quasi-experimental or expost facto. The compensatory program is made available to the most needy, and the "control" group then sought from among the untreated children of the same community. Often this untreated population is on the average more able than the "experimental" group. In such a situation the usual procedures of selection, adjustment, and analysis produce systematic biases in the direction of making the compensatory program look deleterious. Not only does matching produce regression artifacts in this direction, but so does analysis of covariance and partial correlation. This essay illustrates with a detailed example why these biases appear.

Fennessey, James. Blending Evidence, Technique, and Judgment in Educational Research Inference. Final Report. Baltimore: Dept. of Social Relations, Johns Hopkins Univ., November 1976. 298p. ED 143 675.

This final report of a National Institute of Education project explores Bayesian statistical analysis as a paradigm for educational impact studies, particularly studies on the education of the disadvantaged. The position of the report is that much of what is wrong with educational research can be attributed to the use of an inappropriate model for making inferences. The author presents and discusses advantages of Bayesian inference over the "sampling-theory" framework of inference. Particular attention is paid to the demonstration of advantages of the Bayesian paradigm as a basis for representing knowledge when the data are "weak." Computer programs are provided to facilitate the application of Bayesian analysis to the sort of data most frequently encountered in educational evaluation. Also emphasized is the need for careful and explicit specificaiton of the data-generated model before undertaking data analysis.

Hall, Gene E.; Loucks, Susan F. A Developmental Model for Determining Whether the Treatment is Actually Implemented. American Educational Research Journal, v14 n3 p263-276, Summer 1977. EJ 180 513.

Determining whether or not the treatment or innovation under study is actually in use and, if so, how it is being used, is essential to the interpretation of any study. The concept of Levels of Use of the Innovation (LOU) permits an operational, cost-feasible description and documentation of whether or not an educational innovation or treatment is being implemented. Eight different LOUs can be reliably measured: nonuse, orientation, preparation, mechanical uses, routine, refinement, integration, and renewal.

Kennedy, Marcy M. Generalization of Findings from Single Case Studies. Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978. 26p. ED 166 222. (Available in microfiche only.)

Although single case studies might be useful to evaluators for a variety of purposes, there are no generally accepted ways for drawing inferences about the generality of findings from a case study. Single case studies are defined in this paper as either studies of single events, or disaggregated studies of multiple events. The data may be qualitative or quantitative, and may be derived from controlled experiments or from observation. There are two spans to the bridge of inference. The statistical span connects the experimental sample to a population just like that sample. The second span connects the population to a group judged to be sufficiently similar. In case law or in clinical practice, the judgment of sufficient similarity--that is, the judgment of the appropriateness of the generalization--is made by the user. This application of single case data may also be appropriate in educational evaluation.

Lindvall, C. Mauritz. Basic Considerations in Assessing the Validity of Evaluation Designs. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979. 24p. ED 170 369.

Evaluation studies on educational questions attempt to provide answers in the form of conclusions or inferences which are derived from the information collected. Valid inferences are a result of careful research design for the study, and may be causal, descriptive, value-oriented, or probabilistic. Basic steps in designing an evaluation study are suggested: (1) developing examples of each type of inference which will result; (2)identifying major components of each inference; (3) specifying the validity concerns of each component; and (4) planning the study so that inferences will be defensible against claims of invalidity. Types of validity applicable to evaluation studies include content validity, construct validity, internal validity, population validity, and ecological validity.

Porter, Andrew C.; And Others. Practical Significance in Program Evaluation. American Educational Research Journal, v15 n4 p529-539, Fall 1978. EJ 200 567.

Defining practical significance in program evaluations is a difficult measurement problem, which can only be solved by an intimate familiarity with the measures upon which effects are estimated, and their substantive relationship with the goals of the program being evaluated. Past attempts to describe the "size of effect" of instructional programs have characteristically relied on statistical indices that can be estimated and reported without any knowledge of what was measured. This practice is shown to be misdirected. Instead, what is called for is a procedure whereby the substantive instructional intentions of the program, the substantive characteristics of a test, and the interrelationship between the two are made explicit.

D. ACCURACY STANDARDS

11. Objective Reporting. The evaluation procedures should provide safeguards to protect the evaluation findings and reports against distortion by the personal feelings and biases of any party to the evaluation.

Note: See also the references for Standards A2, A4, B2, C2, C3, C7, D2, D4, D5, D6, and D10.

Datta, Lois-ellen. Communicating Evaluation Results for Policy Decision Making. In Berk, Ronald A. (Ed.), Educational Evaluation Methodology: The State of the Art. Baltimore: Johns Hopkins Univ. Press, 1981. Chapter 6.

This chapter reviews techniques and issues in communicating evaluation findings to decision-makers. The literature includes studies of techniques such as adversary evaluation, reports of stakeholder participation studies, and case studies of evaluation utilization. Actual evaluation reports from ten large city school districts and five states are critiqued; the majority are characterized as descriptive statistical accounts rather than action-oriented evaluations. It is argued that effective communication must come at the beginning of an evaluation study as an integral part of the planning and execution.

GENERAL MONOGRAPHS AND TEXTBOOKS

Anderson, Scarvia B.; Ball, Samuel. The Profession and Practice of Program Evaluation. San Francisco: Jossey-Bass, 1978.

Beginning with an overview of the field, this book treats the following topics: the major purposes of evaluating educational and social programs; general methods of evaluation best suited for each purpose; types and sources of evidence frequently associated with the general methods of investigation; targeted dissemination of evaluation information and results; the professional predispositions and preferences of evaluators that may influence what they look at and how they look at it; the complex fiscal and administrative relationships among funding agencies, program directors, and evaluators; ethical responsibilities bound up in program evaluation; defining, instilling, and assessing the competencies of evaluators; and the status and prospects for evaluation as a "profession."

Apple, Michael W.; And Others (Eds.). Educational Evaluation: Analysis and Responsibility. Berkeley, Calif.: McCutchan Pub. Co., 1974.

This volume is concerned with three problems of major significance. First, it aims at broadening the perspectives of educators on the problems of and approaches to evaluation, going beyond the restrictive input-output models that dominate the topic. Secondly, it is aimed at overcoming the limited range of conceptual and historical insights in the evaluation field. Finally, it portrays the intense controversy that any serious discussion of evaluation is bound to raise. After two introductory chapters, six chapters on the concepts, values and methods in evaluation are each followed by a critique and discussion.

Babbie, Earl R. Survey Research Methods. Belmont, Calif.: Wadsworth Pub., 1973.

This book is addressed to three problems related to the misconception that survey research is simple. First, the faddish popularity of survey methods has inevitably resulted in a large number of bad surveys. Second, the widespread overuse and misuse of survey methods has led to the wholesale rejection of survey research by many people. Third, the assertion that a given survey was poorly conducted incorrectly presupposes an established body of scientific standards against which to evaluate survey activities. This text focuses on the logic and skills of survey research, in an attempt to provide a practical guide to survey research for students and other prospective researchers.

Baker, Eva L.; Quellmalz, Edys S. (Eds.) Educational Testing and Evaluation: Design, Analysis, and Policy. Beverly Hills, Calif.: SAGE Pub., 1980.

The papers collected here were presented at an invitational conference on Measurement and Methodology in Education, sponsored by the Center for the Study of Evaluation. In general, the participants paid relatively little attention to educational policy as a legitimate focus for activity; nor do they concretely express concern for students, those who are the data providers in most testing and evaluation efforts. They do directly address issues in test design, quantitative theory and applications, and evaluation and testing policy.

Banner, David K.; And Others. The Politics of Social Program Evaluation. Cambridge, Mass.: Ballinger Pub. Co., 1975.

Attempts to develop evaluation procedures for social programs have been fraught with difficulties, obstacles, and political barriers. These are described in this book in the context of the early history of the Opportunity Funding Corporation (OFC), a demonstration activity established and funded by the Office of Economic Opportunity during the first Nixon administration. This case study relies primarily on personal interviews and selected secondary sources; the data were then analyzed in light of relevant theory on the politics of evaluation.

Bennett, Carl A.; Lumsdaine, Arthur A. (Eds.) Evaluation and Experiment: Some Critical Issues in Assessing Social Programs. New York: Academic Press, 1975.

This volume is an outgrowth of a symposium held at the Battelle Seattle Research Center in July, 1973. It focuses on some selected aspects of the problems in evaluating the outcomes of socially important programs. Its eight chapters cover the defining of evaluation issues; assessment as an empirical base for policy; effect size estimation in quasi-experimental designs; regression and selection models to improve nonexperimental comparisons; field trial designs in gauging the impact of fertility planning programs; a reexamination of experiments and evaluations; operational and systematic research on production, maintenance, control, and adaptive functions of feedback; and assessing alternative conceptions of evaluation.

Borich, Gary D. (Ed.) Evaluating Educational Programs and Products. Englewood Cliffs, N.J.: Educational Technology Pub., 1974.

This book is a guide and handbook for planners, developers, and evaluators of educational programs and products. Its purpose is to provide practical insights that are immediately applicable to planning and executing effective program and product evaluations. It divides the evaluator's work into three activities: establishing perspective, or choosing an appropriate role for the context in which he or she will work; planning the evaluation, or choosing an appropriate model or strategy; and analyzing the data, or selecting appropriate analysis methods and techniques. The task of this book is to identify specific procedures that are appropriate to each of these activities.

Campbell, Donald T.; Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally & Co., 1963.

The validity of 16 experimental designs against 12 common threats to valid inference is examined. The designs examined are the one-shot case study, the one-group pretest-posttest design, the static-group comparison, the pretest-posttest control group design, the Solomon four-group design, the posttest-only control group design, the time-series experiment, the equivalent time-samples design, the equivalent materials design, the nonequivalent control group design, counter balanced designs, the separate-sample pretest-posttest design, the separate-sample pretest-posttest control group design, the multiple time-series design, the recurrent institutional cycle design, and regression-discontinuity analysis.

Caro, Francis G. (Ed.) Readings in Evaluation Research. Second Edition. New York: Russell Sage Foundation, 1977.

This volume brings together material about evaluation research drawn from a variety of sources, and includes both general statements about evaluation research and specific case materials. The general articles address such issues as the nature of the evaluation task, the role of evaluation research in programs of directed change, the organizational context in which evaluation research is conducted, and the methodological strategies appropriate for evaluation research. The case materials include treatment of problems in the establishment of the evaluation research role and reports of findings of completed evaluation research studies.

Cook, Thomas D.; And Others (Eds.) Evaluation Studies Review Annual. Volume 3. Beverly Hills, Calif.: SAGE Pub., 1978.

The third volume in this series includes papers on the policy and political context of evaluation, methodology, and exemplary studies from the fields of health, income maintenance, criminal justice, education, mental health, and evaluations in the "public interest."

Cook, Thomas D.; Reichardt, Charles S. (Eds.) Qualitative and Quantitative Methods in Evaluation Research. Beverly Hills, Calif.: SAGE Pub., 1979.

A diversity of opinion on the use of qualitative and quantitative methods was purposely included in this volume. William J. Filstead distinguishes between the qualitative and quantitative paradigms, and argues that the qualitative paridigm is most appropriate for evaluation research. Donald T. Campbell provides a convincing rationale for the use of qualitative methods in the case study design. M. G. Trend describes an evaluation of a major housing allowance demonstration which used both participant observers and questionnaire surveys. Francis A. J. Ianni and Margaret Terry Orr argue that evaluators can profitably use ethnographic techniques only if these methods are carefully adapted to fit the needs of evaluation research. Howard S. Becker discusses the "truth" of photographic evidence. Michael S. Knapp provides an examination of the use of ethnography in the evaluation of the Experimental Schools Program. Robinson G. Hollister, Peter Kemper, and Judith Wooldridge offer some insights into the use of quantitative methods.

Cronbach, Lee J.; And Others. Toward Reform of Program Evaluation. San Francisco: Jossey-Bass, 1980.

In an attempt to bring about change in thought by provoking argument, the main ideas of the book are first presented as 95 theses. These theses include statements about the role of evaluation; the political impact of an evaluation; the usefulness of evaluation for decision-making; the information dissemination role of evaluation; the relation between program goals and evaluation; the design of an evaluation; evaluation validity; evaluation as a profession; and the various roles of the evaluator.

Cronbach, Lee J. Essentials of Psychological Testing. Third Edition. New York: Harper & Row, 1970.

This book is intended to establish a base of the essentials of measurement: methods of inquiry, the critical standards, and the key concepts of the field. Its chapters address the purposes and types of tests; test administration, scoring, and validation; other characteristics desired in tests; tests of ability; interest inventories; and personality measures.

Datta, Lois-ellin; Perloff, Robert (Eds.) Improving Evaluations. Beverly Hills, Calif.: SAGE Pub., 1979.

Dedicated to Marcia Guttentag, the four sections of this volume reflect interests prominent in her published work. The first section deals with the political forces influencing what questions are asked, when, and by whom. Writers in the second section share a common concern with the question of causal inference or internal validity: how to feel reasonably assured that effects, if any, are attributable to the change under study. The third section deals with three measurement concerns. The final section returns to discussions of the relation of evaluation to the social order in which it is inextricably embedded.

Freeman, Howard E.; Solomon, Marian A. (Eds.) Evaluation Studies Review Annual. Volume 6. Beverly Hills, Calif.: SAGE Pub., 1981.

The sixth volume in this series includes papers on concepts and approaches, and evaluation utilization, and exemplary studies from the fields of education, human resources and social services, law and public safety, health, mental health and substance use, and environment.

Glass, Gene V. (Ed.) Evaluation Studies Review Annual. Volume 1. Beverly Hills, Calif.: SAGE Pub., 1976.

The first volume in this series includes papers on the theory and methods of evaluation, and exemplary studies from the fields of education, mental health and public health services, welfare and social services, and crime and justice.

Glass, Gene V.; Stanley, Julian C. Statistical Methods in Education and Psychology. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

This textbook is for use in statistics courses in education and the social sciences. Its chapters cover measurement, scales, and statistics; tabulating and depicting data; measures of central tendency; measures of variability; the normal distribution; measures of relationship; prediction and estimation; probability; statistical inference; analysis of variances; multiple comparison procedures; and fundamentals of experimental design.

Grotelueschen, Arden D.; And Others. An Evaluation Planner: A Guidebook for Developing Evaluation Plans Responsive to a Variety of Contexts, Audiences, and Issues Within Adult Basic Education. Urbana: Univ. of Illinois, Office for the Study of Continuing Professional Education, 1974.

Systematic guidelines for evaluating programs in adult basic education are presented. Program activities in adult basic education involve four main levels of activity (classroom, local, state, and federal), each of which can profit from appropriately focused evaluation. Patterns of association among personnel and activities are present in each educational context, as well as common evaluation needs and emphasis. Typical characteristics of each context are described. Eight evaluation components form the basis of the planner: purposes, audiences, issues, resources, evidence, data-gathering, analysis, and reporting. A discussion of each component is followed by a checklist containing suggestions, possibilities, and techniques in developing an evaluation design for each of the four contexts.

Guba, Egon G.; Lincoln, Yvonna S. Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches. San Francisco: Jossey-Bass, 1981.

This book presents a new model of evaluation--one that organizes evaluation activities so that it illuminates the claims, concerns and issues raised by stakeholding audiences (responsive evaluation) and uses naturalistic methodologies to gather information. A variety of evaluation models are analyzed to provide background information necessary to an understanding of the responsive approach. Scientific and naturalistic inquiry paradigms are compared and a case is made that the latter be used in the study of human behavior. Problems associated with using a human being as an assessment instrument are discussed as well as ways of improving the human instrument. There is also a series of chapters discussing the methods and methodologies associated with qualitative inquiry. The last part of the book is a presentation of steps by which naturalistic--responsive evaluation is carried out. Among the items discussed are contracting for an evaluation, establishing the evaluators presence on the site, developing contacts, avoiding overinvolvement and cooptation and dealing with human and political problems. There is also discussion on how to deal with audiences and elicit their claims, concerns and issues. The methods of identifying the kinds and sources of useful information are presented. The book concludes with a chapter on reporting, emphasizing different audiences, different reporting requirements, and that reporting is continuous.

Guttentag, Marcia; Saar, Shalom (Eds.) Evaluation Studies Review Annual. Volume 2. Beverly Hills, Calif.: SAGE Pub., 1977.

The second volume in this series includes papers on thinking about evaluation, evaluation methodology and data integration, evaluation into policy, and exemplary studies from the fields of education, crime and justice, and human services.

Guttentag, Marcia; Struening, Elmer L. (Eds.) Handbook of Evaluation Research. Two volumes. Beverly Hills, Calif.: SAGE Pub., 1975.

One purpose of this Handbook is to provide evaluators with a sample of experts, a panel of consultants between hard covers, with whom they can communicate as they develop the crucial steps of their studies. The Handbook is also designed to serve as a textbook for courses in evaluation at the graduate level. The first volume emphasizes strategies and methods of evaluation, while the second volume reviews the literature in selected content areas.

Hamilton, David; And Others (Eds.) Beyond the Numbers Game: A Reader in Educational Evaluation. London: Macmillan Education, 1977.

This book charts the paradigm shift from an evaluation methodology valuing numeracy to one valuing literacy. The many readings are organized into four sections: "The Objectives Model Revisited," "Five Advocates of Change" (Myron Atkin, Elliot Eisner, Lawrence Stenhouse, Michael Scriven, and Robert Stake), "Alternative Methodology," and "Alternative Evaluation: The New Paradigm in Action."

Hays, William L. Statistics for Psychologists. Second Edition. New York: Holt, Rinehart and Winston, 1973.

This book represents an attempt to give the elements of modern statistics in a relatively nonmathematical form, but in somewhat more detail than is customary in texts designed for psychologists, and with considerably more emphasis on the theoretical rather than the applied aspects of the subject. Its chapters address sets and functions, elementary probability theory, the binomial distribution, central tendency and variability, sampling distributions and point estimation, the normal distribution, hypothesis testing, inferences about population means, the chi-square and F distributions, analysis of variance, and linear regression and correlation.

Heise, David R. Causal Analysis. New York: John Wiley & Sons, 1975.

This book focuses on the study of linear systems and represents an effort to organize a broad range of information about this topic in a fairly elementary fashion. Its chapters consider causality and causal analysis, causal diagrams and flowgraph analysis, statistical concepts, path analysis, identification and estimation, and dynamic considerations.

House, Ernest R. (Ed.) School Evaluation: The Politics & Process. Berkeley, Calif.: McCutchan Pub. Co., 1973.

The primary purpose of this book is to acquaint the practicing administrator, the decision-maker, and the educational-consumer with the world of evaluation. Its major theme is the political nature of evaluation. A second theme is that both one's administration and one's evaluation are intimately related to whether one believes that the goals of the class, school district, state, or nation are already established or are yet to be arrrived at through negotiation between groups. A third theme is the relationship between decision-making and evaluation: the delicate relationship between the administrator and the evaluator.

Morris, Lynn Lyons; And Others. Program Evaluation Kit. Beverly Hills, Calif.: SAGE Pub., 1978.

The Kit is a set of books intended to assist people who are conducting evaluations of educational programs. It consists of the following eight books: The Evaluator's Handbook; How to Deal With Goals and Objectives; How to Design a Program Evaluation; How to Measure Program Implementation; How to Measure Attitudes; How to Measure Achievement; How to Calculate Statistics; and How to Present an Evaluation Report.

Patton, Michael Quinn. Qualitative Evaluation Methods. Beverly Hills, Calif.: SAGE Pub., 1980.

This book is designed to explain in which evaluation situations qualitative methods are useful, and how to actually use those methods. It is organized around three topics: conceptual issues in the use of qualitative methods for evaluation research, collecting qualitative data, and data analysis.

Popham, W. James. Educational Evaluation. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

This textbook is designed to be appropriate for beginning educational-evaluators, and graduate students in educational evaluation. Its chapters address contemporary conceptions of evaluation, instructional objectives, the use of measurement, expanding measurement alternatives, classical measurement considerations, criterion-referenced measurement, measurement of affect, evaluation designs, sampling strategies, analyzing evaluative data, reporting evaluation results, cost analysis considerations, and teacher evaluation.

Popham, W. James (Ed.) Evaluation in Education: Current Applications. Berkeley, Calif.: McCutchan Pub. Co., 1974.

This volume was initiated by the Standing Committee on Research Training of the American Educational Research Association. Its chapters include: "Evaluation Perspectives and Procedures," by Michael Scriven; "Alternative Approaches to Educational Evaluation: A Self-Study Guide for Evaluators," by David L. Stufflebeam; "Designing Summative Evaluation Studies at the Local Level," by Peter W. Airasian; "Data Analysis and Reporting Considerations in Evaluation," by Richard M. Wolf; "The Use of Standardized Tests in Evaluation," by Gilbert Sax; "Criterion-Referenced Measurement," by Jason Millman; "Cost Analysis for Educational Program Evaluation," by Emil J. Haller; "Introduction to Matrix Sampling for the Practitioner," by Kenneth A. Sirotnik; and "Formative Evaluation of Instruction," by Eva L. Baker.

Rosenthal, Robert; Rosnow, Ralph L. (Eds.) Artifact in Behavioral Research. New York: Academic Press, 1969.

That portion of the complexity of human behavior which can be attributed to the social nature of behavioral research can be conceptualized as a set of artifacts to be isolated, measured, considered and, sometimes, eliminated. This book is designed to consider in detail a number of these artifacts. The introductory chapter provides a perspective on artifact and a discussion of the nature of experimental control. The following six chapters are a series of position papers by researchers who have been actively engaged in systematic exploration of various antecedents of artifact in behavioral research, and each writer summarizes the findings in his respective area: suspiciousness of intent, volunteer effects, pretest sensitization, demand characteristics, experimenter expectancy effects, and evaluation apprehension. The final chapter takes into account the separate contributions and discusses the future prospects for behavioral research.

Rossi, Peter H.; Williams, Walter (Eds.) Evaluating Social Programs: Theory, Practice, and Politics. New York: Seminar Press, 1972.

This volume of readings is divided into four sections: An Overview; Evaluative Research: Theory; Evaluative Research: Practice; and Organizing for Large-Scale Evaluative Research. The papers in the Theory section address statistical design requirements and methodological issues. The papers in the Practice section illustrate evaluation in education, labor programs, and income maintenance experiments. The final section discusses the roles in evaluation research of various social institutions.

Sechrest, Lee; And Others (Eds.) Evaluation Studies Review Annual. Volume 4. Beverly Hills, Calif.: SAGE Pub., 1979.

The fourth volume in this series includes papers on the theory and philosophy of evaluation, alternative methodologies and strategies, the technology of evaluation, unanticipated findings, evaluation utilization, and a variety of evaluation studies.

Sechrest, Lee (Ed.) Unobtrusive Measurement Today. New Directions for Methodology of Behavioral Science, n1, 1979.

The focus of this volume is on the appropriate utilization of unobtrusive measures in research. An overview addresses some of the critical methodological and conceptual issues that faces the researcher wishing to devise and employ an unobtrusive measure. Other chapters include discussions on direct observation, application in field experiments and cross-cultural research, nonverbal behaviors, and the physical trace.

Stromsdorfer, Ernst W.; Farkas, George (Eds.) Evaluation Studies Review Annual. Volume 5. Beverly Hills, Calif.: SAGE Pub., 1980.

The fifth volume in this series includes papers on methodology, and exemplary designed and natural experiments from the fields of labor, education, housing, health and safety, energy and resources, and public financial policy.

Thorndike, Robert L.; Hagen, Elizabeth P. Measurement and Evaluation in Psychology and Education. Fourth Edition. New York: John Wiley & Sons, 1977.

The chapters of this measurement textbook address: measurement and numbers; qualities desired in a measurement procedure; norms; various types of tests and measurement instruments; planning a school testing program; grading; and social and political issues in testing.

Tuckman, Bruce W. Conducting Educational Research. New York: Harcourt Brace Jovanovich, 1972.

This book is based on the premises that research is a useful tool for educators, and that much educational research must be undertaken in the field. It discusses the role of research, selecting a problem and constructing hypothesis, identifying and labeling variables, constructing operational definitions of variables, identifying techniques for the manipulation and control of variables, and constructing research designs.

Tyler, Ralph W.; And Others (Eds.) Perspectives of Curriculum Evaluation. Chicago: Rand McNally & Co., 1967.

This volume is the first in a Monograph Series on Curriculum Evaluation, developed by the American Educational Research Association. Its chapters include: "Toward a Technology for the Evaluation of Educational Programs," by Robert E. Stake; "Changing Concepts of Educational Evaluation," by Ralph W. Tyler; "Curriculum Research and the Promotion of Learning," by Robert M. Gagne; "The Methodology of Evaluation," by Michael Scriven; and "Aspects of Curriculum Evaluation: A Synopsis," by J. Stanley Ahmann.

Weiss, Carol H. Evaluating Action Programs: Readings in Social Action and Education. Boston: Allyn and Bacon, 1972.

This book aims to help the reader conceptualize and understand the purposes of evaluation and the methods by which it obtains information and generates conclusions. Rather than giving a set of prefabricated rules and instructions, it points out the constraints within which evaluation operates and suggests alternative strategies of design, measurement, structure, relationship, and communication in order to accommodate to existing constraints and to serve the informational needs of programs.

Weiss, Carol H. Evaluation Research: Methods for Assessing Program Effectiveness. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

This book deals with the application of research methods to the evaluation of social programs: programs in education, social work, corrections, health, mental health, job training, technical assistance, community action, and law. Its basic theme is that evaluation uses the methods and tools of social research but applies them in an action context that is intrinsically inhospitable to them. Its chapters discuss the purposes of evaluation, formulating the question and measuring the answer, design of the evaluation, the turbulent setting of the action program, and utilization of evaluation results.

Worthen, Blaine R.; Sanders, James R. Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles A. Jones Pub. Co., 1973.

This book pulls together in one volume the best of the emerging literature on educational evaluation, and identifies and fills serious gaps in the literature. It is organized around four topics: evaluation as disciplined inquiry, frameworks for planning evaluation studies, considerations in planning evaluation studies, and the future of evaluation.

INDEX

Agarwala-Rogers, Rehka, A8
Ahn, Unhai R., A2
Alderman, Donald L., B1 (Powers)
Alexander, Jay, A8
Alternative Methodology, D9
Anderson, Scarvia B., A4, A6 (Ball), A7, Monograph
Apple, Michael W., A4, Monograph
Baker, Eva L., Monograph
Ball, Samuel, A4 (Anderson), A6, D7, Monograph (Anderson)
Banner, David K., B2, Monograph
Baumrind, Diana, C5
Becker, Howard S., D9
Beers, C. David, C2 (Clark)
Bennett, Carl A., Monograph
Bentler, Peter M., D8
Berk, Ronald A., A4, D5
Berlak, Harold, A4
Bogdan, Robert, D9
Borich, Gary D., Monograph
Boruch, Robert F., B1
Brager, Gary L., A5
Braskamp, Larry A., A2, A8 (Brown)
Brickell, Henry M., B2
Brown, Robert D., A5, A8
Bryk, Anthony S., D8
Burges, Bill, D9 (Ferreira)
Campbell, Donald T., D4, D5 (Cook), D6 (Cook), D10, Monograph
Caro, Francis G., B2, Monograph
Casper, Paul N., B1
Chibucos, Thomas R., D8 (Porter)
Ciarlo, James A., A8
Clark, Woodrow W., Jr., C2
Cochran, Nancy, D4
Cook, Thomas D., D5, D6, Monograph (2 citations)
Cooley, William W., D8
Cost of Educational Accountability, B3
Cox, G., A8
Craig, Marilyn Martin, A3
Cronbach, Lee J. D5 (2 citations) D6, D8 (2 citations), Monograph (2 citations)
Datta, Lois-ellen, D11, Monograph
David, Jane L., D4
Davis, Howard R., A8
Demaline, Randy E., D7
Denny, Terry, D2
Durham, Robert L., D5 (Nunnally)
Eichelberger, R. Tony, A4
Eignor, Daniel R., D5 (Hambleton)
Englert, Richard M., B2
Erlebacher, Albert, D10 (Campbell)
Ethical Standards of Psychologists, C5
Everhart, Robert B., C6, D9
Farkas, George, Monograph (Stomsdorfer)
Fennessey, James, D10
Ferreira, Joseph, D9
Fienberg, Stephen E., D9
Filstead, William J., D9
Final Regulations, C5
Freeman, Howard E., Monograph
Fullan, Michael, D1
Furby, Lita, D8 (Cronbach)
Glass, Gene V., Monograph (2 citations)
Goodrich, Thelma Jean, A4 (Gorry)
Gooler, Dennis D., C4
Gorry, G.Anthony, A4
Granville, Arthur C., A8
Green, Donald Ross, D5 (Wargo)
Grobe, Robert P., A8
Gross, Alan L., A4
Grotelueschen, Arden D., Monograph
Guba, Egon G., A4 (Lincoln), A8, D9, Monograph
Gurel, Lee, A2
Guttentag, Marcia, A4, Monograph (2 citations)
Haenn, Joseph F., A8
Hagen, Elizabeth P., Monograph (Thorndike)
Hall, Gene E., D10
Hambleton, Ronald K., D5
Hamilton, David, Monograph
Hayman, John, A3
Hays, William L., Monograph
Heise, David R., Monograph
Hess, Robert J., A1
House, Ernest R., A1, A4 (3 citations), A5, B2, C1, Monograph

Johnson, Mauritz, A4
Joyce, John F., C6
Kelman, Herbert C., C5
Kennedy, Marcy M., D10
Kenny, David A., D8
Kolodny, Ralph L., C6 (Rodman)
Krathwohl, David R., A4 (2 citations)
Krippendorff, Klaus, D9
Kunkel, Richard C., A4
Leinhardt, Gaea, D1
Leithwood, Kenneth A., D1
Levine, Adeline, B2
Levine, Murray, B2
Likert, Rensis, A8 (Mann)
Lincoln, Yvonna S., A4, Monograph (Guba)
Lindvall, C. Mauritz, D10
Linn, Robert L., D6
Lofland, John, D9
Lord, Frederic M., D6
Loucks, Susan F., D10 (Hall)
Lumsdaine, Arthur A., Monograph (Bennett)
Mager, R. F., D3
Mann, Floyd, A8
Marascuilo, Leonard A., D8
Mathis, William, B2
Mazza, Paul, A5 (Brager)
McGranahan, Pamela, A1
Messick, Samuel, A4, D5
Metfessel, Newton S., A3
Michael, John A., C5 (Weinberger, 2 citations)
Michael, William B., A3 (Metfessel)
Millman, Jason, A2
Molner, Stanley F., C2
Montgomery, Deborah J., D1 (Leithwood)
Morris, Lynn Lyons, Monograph
Murphy, Richard T., D7
Myrdal, Gunnar, A4
Nafziger, Dean H., D3 (Sanders)
Neigher, William, C2 (Windle)
Newman, Dianna L., A2
Newman, Warren B., A2
Novak, Carl D., A8
Nunnally, Jum C., D5, D6
Ott, Jack M., A3
Patton, Michael Q., A1, A6, A8, D1, Monograph
Perloff, Robert, Monograph (Datta)
Polemeni, Anthony J., B2
Pomfret, Alan, D1 (Fullan)
Popham, W. James, A5, Monograph (2 citations)
Porter, Andrew C., D5, D8, D10
Powers, Donald E., B1
Poynor, Hugh, D4
Preparing Evaluation Reports, C7
Quellmalz, Edy S., Monograph (Baker)
Quinn, D. William, D7
Reichardt, Charles S., Monograph (Cook)
Relavin, Sol H., D4 (David)
Roberts, Sarah, A5
Rodman, Hyman, C6
Roecks, Alan L., B1 (Casper)
Rose-Pendleton, M. K., C1
Rosenthal, Robert, Monograph
Rosnow, Ralph L., Monograph (Rosenthal)
Rossi, Peter H., A4 (Berk), Monograph
Saar, Shalom, Monograph (Guttentag)
Salasin, Susan E., A8 (Davis)
Sanders, James R., A2, D3, Monograph (Worthen)
Sanders, Nancy, C2 (Sieber)
Sawin, Enoch I., D4
Scheirer, Mary Ann, C2
Scheyer, Patricia T., B1, D2
Schriber, Peter E., B3
Scriven, Michael, A4, C2
Sechrest, Lee, Monograph (2 citations)
Sheinfeld, Sherri Nita, C2
Shepard, Lorrie, D5
Shiffer, Lois J., C5
Sieber, Joan E., C2
Sjoberg, Gideon, A4
Sjogren, Douglas D., D1
Sladek, Frea E., C8
Slinde, Jeffrey A., D6 (Linn)
Smith, Louis M., D9 (2 citations)
Smith, Nick L., A4
Solomon, Marian A., Monograph (Freeman)

Sroufe, Gerald E., B2
Staff Response, A, C1
Stake, Robert E., C3, D2
(Scheyer), D4
Standards for Educational
and Psychological Tests, D5
Stanley, Julian C., D6,
Monograph (Campbell),
Monograph (Glass)
Stein, Eugene L., C8 (Sladek)
Steinmetz, Andres, D1
Stevens, W. F., A6
Straton, Ralph G., A1
Stromsdorfer, Ernst W.,
Monograph
Struening, Elmer L,. Monograph
(Guttentag)
Stufflebeam, Daniel L., C1, D3, D7
Taylor, Steven J., D9 (Bogdan)
Thorndike, Robert L., Monograph
Tittle, Carol Kehr, A8 (2 citations)
D4 (2 citations)
Tornatzky, L. G., A6 (Stevens)
Trend, M. G., D9
Tucker, Susan A., A4 (Kunkel)
Tuckman, Bruce W., Monograph
Tumin, Melvin M., B2
Tyler, Ralph W., Monograph
Ulschak, Francis L., C6
Walker, Clinton B., D5
Wargo, Michael J., D5
Weinberger, JoAnn, C5 (2 citations)
Weiner, Stephen S., C1
Weisberg, Herbert I., D8, D8 (Bryk)
Weiss, Carol H., A8, B2, C6,
Monograph (2 citations)
Weiss, Roland G., C6, (Ulschak)
Wick, John W., B1
Williams, Walter, Monograph
(Rossi)
Windle, Charles, C2
Wolcott, Harry, D9
W f, Richard M., A5, D8
Woodward, J. Arthur, D8
(Bentler)
Worthen, Blaine R., C1
(Wright, 2 citations),
Monograph
Wright, William J., A1 (Hess),
B2, C1 (2 citations)